

THIS WEEK

EDITORIALS

ARCTIC Mass walrus beaching
a sign of warmer things to
come **p.140**

WORLD VIEW European
scientists unite to protest
against cuts **p.141**

BUGS LIFE Nasal bacteria
switch genes to arm for
invasion **p.143**

Out of Africa

The Ebola outbreak in West Africa must be shut down now, or the disease will continue to spread.

Ebola is out of control in Liberia, Sierra Leone and Guinea. Although this has been the case since late spring, the international pledges of help have yet to translate into concerted, rapid action on the ground. The virus still has the upper hand. Between 23 September and 1 October alone, the number of cases rose from 6,500 to almost 7,500, according to the World Health Organization.

The situation has become so bad that no one knows the true numbers of cases and deaths, only that they are probably much larger than official estimates. And the accompanying collapse of the countries' health systems means that people are unnecessarily falling victim to malaria and other diseases. One does not need a mathematical model to foresee how bad this could get.

The biggest risk is that the outbreak will spread to neighbouring countries. The risk to the rest of the world, and in particular to richer countries, remains low. But there is no room for complacency. Spain and the United States have now had cases, and the hapless management of the US case serves as a reminder that even a country with probably the most developed plans for responding to biological threats — be they natural or terrorist — can fall far short when put to the test. The infected man had flown from Liberia to Dallas, Texas, on 20 September. Four days later, he started to develop symptoms; he went to the hospital on 26 September, but no red flags were raised and he was sent home. He was not isolated until 28 September, after returning to the hospital in an ambulance.

The people he had been staying with were then quarantined, but spent more than a week in an apartment that still contained his Ebola-contaminated clothing and linen, which they had to decontaminate and bag as best they could. Meanwhile, local, state and federal officials argued over their disparate laws governing the decontamination and transport of dangerous pathogens, before belatedly sending in a commercial clean-up team.

The US health system is sufficiently robust to stop a full-blown outbreak developing. But even in rich countries, inequalities in access to health care and cost-cutting in the health services can create vulnerabilities. Strong public-health systems, with sufficient slack in them to provide the surge capacity that is needed in a serious outbreak, are a crucial defence.

The average life expectancy in the United States lags behind that of many other developed nations. Although its richest areas boast some of the highest expectancies in the world, some places have expectancies even lower than those in developing countries such as Bangladesh. Were Ebola to spread in underprivileged urban areas, it might not be so easy to control as US officials are making out. The uninsured, in particular, may think twice about going to see a doctor, and so hamper efforts to stem an outbreak.

The media's handling of the first US-diagnosed case also carries lessons. Although much of it has been first-class, the dogged determination to identify the infected man and who he came into contact with could be counterproductive. People who suspect they might have been

in contact with someone infected with Ebola might now be reluctant to come forward in case their names are splashed all over the headlines. The public has a legitimate interest in knowing the places an infected person has frequented, for example, but there is a fine line between this and blatant voyeurism, invasion of privacy and sensationalism.

More broadly, the excessive media and political focus on the threat to the United States and other Western countries — which is relatively low — risks engendering a siege mentality. And as Thomas Frieden, director of the US Centres for Disease Control and Prevention, has pointed out, flight bans and other restrictive measures only hamper access by relief agencies, and diminish the much-needed cooperation of local communities. The first US-diagnosed case, he says, is a stark reminder that

the threat of exported cases will persist until the outbreak is quashed.

The world is fiddling while West Africa burns, and unless it acts much faster, the outbreak risks spreading to surrounding regions. Sparks from it could lead to exports to more far-flung places, perhaps even to major cities that lack decent public-health infrastructure. But countries and the public must also realize that although action is needed urgently, the commitments must be sustained until the outbreak has been stamped out, which could take many months. The relatively low threat to developed countries must not distract or detract from the pressing need to tackle the outbreak at its source. ■

**“The world is
fiddling while
West Africa
burns.”**

A little knowledge

The significance of expertise passed on by direct contact — tacit knowledge — is moot.

For the last two decades of the twentieth century, a cold war rumbled on between the laboratories of physicists in Moscow and in the West over the quality of sapphire. The Russian scientists claimed to have measured the rate of decay of the material's resonance — a signal of its quality — with what researchers elsewhere considered impossible precision. The stakes were high: sapphire mirrors were being considered for use in a new generation of laser interferometer gravitational-wave detectors. But were they up to the task? Labs in the United States and United Kingdom could not reproduce the Moscow findings. The discrepancy fuelled mistrust and antagonism.

At the turn of the millennium, the mystery was solved. Measuring the quality of sapphire, it turns out, is as much art as science. The Moscow scientists were expert experimenters, but this expertise was not transferred through the methods sections of their academic papers. The fine fibres used to suspend the sapphire cylinders under investigation

were greased with “the presence of a fatty film”, one of their translated papers pointed out. Less explicit was the source of the grease. Only after years of struggling with various lubricants did the Western researchers realize that one member of the Russian group would sometimes run the thread across the bridge of his nose or behind his ear. With the right amount of human ‘flossing’ (and the right human), the Western scientists managed to get similar results.

The thread greasing is an example of tacit knowledge: know-how that can be passed on only through direct contact, and not by written or verbal instruction. How to ride a bicycle is a classic case. How to make an atomic bomb is a less-well-known example: all the instructions to build a nuclear weapon may be there on the Internet, but the ‘been there, done that’ personal experience is not. Indeed, security analysts have suggested that the lack of active testing and consequent erosion of nuclear-weapon tacit knowledge is leading to the “uninvention” of the bomb, and reduced credibility of the nuclear deterrent.

In a paper published this month in the journal *Science and Public Policy*, researchers in the United Kingdom suggest that a reverse process is under way when it comes to biology and biological weapons (J. Revill and C. Jefferson *Sci. Public Policy* 41, 597–610; 2014). Access to tacit knowledge in the life sciences is not dwindling but proliferating, argue James Revill and Catherine Jefferson. As secrets are shared, chiefly through advances in information and communications technology, tacit knowledge becomes explicit and barriers are demolished. And it is worth considering, they point out, what those barriers have held back. Many attempts have been made to manufacture deadly ricin, for example, probably based on Internet recipes, but most fail to truly weaponize the material by not milling it to the necessary particle size.

“Even where optimised weaponisation is not the strategic goal,” the authors write, “tacit knowledge may be an important limiting factor in the ability of unskilled actors to exploit advances in [science and technology], which has important implications for the way in which threat is assessed.” Policy-makers who try to proscribe the

development of biological weapons should first gain a better idea of what is scientifically possible and what is not, they suggest, and a sense of how that is changing.

What has any of this to do with the work of a regular bench scientist? Quite a lot, actually. As more attention is paid to the ‘reproducibility crisis’ in science, and journals and funders wrestle with how to make published research findings more robust, tacit knowledge has emerged as both a problem and an opportunity.

In a Comment piece in this journal last year (M. Bissell *Nature* 503, 333–334; 2013), Mina Bissell warned that the push to replicate findings could unfairly malign research (and researchers) that — just like measuring sapphire quality — relies on tacit-knowledge techniques that are better learnt than studied. The *Journal of Visualized Experiments* already aims to narrow the gap between tacit and explicit knowledge by requiring scientists to video their techniques, and so show colleagues how to conduct procedures, rather than simply telling them.

There are known unknowns and unknown unknowns, as former US defence secretary Donald Rumsfeld clumsily explained. Some tacit knowledge is deliberately withheld, and some journal methods sections offer insufficient space for elaboration. Those are the known unknowns and are most easily addressed. The tacit knowledge that is harder to pass on is the nugget of information that neither the teacher nor the pupil realized was important: the varnish on the Stradivarius violin; the greasing of the thread behind the ear.

The trend in science is towards greater openness and data sharing. Communication is instant and in real time; knowledge has never been more fluid. Science traditionally argues that this is a good thing. There is no inherently good or bad technology, goes the mantra, only good and bad applications. Is the same true for all forms of knowledge? One way or another, we could be poised to find out. ■

“Knowledge has never been more fluid — a good thing, science traditionally argues.”

Holy cows

A mass beaching of walrus in Alaska is a sign of things to come.

For more than a century the central attraction of the Horniman Museum in London has been a too-large stuffed walrus. Victorian taxidermists, the story goes, had never seen a live walrus, so they simply kept filling the floppy hide until the creature seemed to fit its skin. The bloated specimen spends its days looking down on visitors with an erect and noble posture that it never held in life.

Compared to the photogenic polar bear, the walrus, even one as smoothed for the camera as the Horniman’s, makes an unlikely poster species for climate change. But cram the creatures together — 35,000 of them — on a remote Arctic beach, and impose a no-fly zone above to prevent the carnage of a stampede, and it is tempting to see them as the natural world’s latest distress beacon to warn of the creeping chaos of global warming.

In the last week, environmental campaigners have cried that the mass walrus beaching in Alaska, first spotted last month, is another clear signal of our warming world. Climate-change sceptics insist that the event is nothing unusual, and have dug out records of previous mass walrus ‘haul-outs’ to support their case. Delighted by the novelty (images of melting glaciers are so 2009), much of the media has discussed the story, and tried hard to work in a Beatles song reference.

There is a simple way to tell this tale. Walrus spend much of their time out of the water, especially when they are rearing young. They prefer to perch on floating sea ice, which gives them access to the

seabed, where most of their food lives. As sea ice retreats north — and this year provided the sixth-lowest extent of summer Arctic ice on record — more walrus have to haul themselves onto the coast. Since 2000, increasing numbers of Pacific walrus (the Atlantic population is less affected) have been forced onto the beaches around the Chukchi Sea; in October 2010, scientists counted a gathering of 120,000 at Cape Serdtse-Kamen in Russia.

Walrus haul-outs on the coast tend to be dominated by adult males. The current event features substantial numbers of mothers and young, which makes it more worrying. A walrus stampede might sound unlikely, but it is a genuine risk. Spooked animals rush to the water and trample anything in their path. The demographics also suggest that something out of the ordinary is going on: female walrus usually recognize the risks of mass haul-outs, and leave the bulls to it.

The link between increased walrus haul-outs on Arctic beaches and the decreased availability of sea ice is clear-cut. The link to climate change is less so, at least in the short term. Sea-ice cover fluctuates with wind and currents from year to year, and the key for walrus is the position of the ice as much as its extent. They need ice over the continental shelf so that they can both rest and feed. In 2008, when remnant ice remained in the Chukchi Sea in the shallow waters of the shelf, walrus did not come ashore in significant numbers.

Whether or not this particular haul-out of walrus in Alaska is a result of climate change is ultimately a moot point. Annual peaks and troughs — of animal movement and ice measurements — are symbolic, but the long-term trend is clear: the Arctic is warming, its ice is melting and the walrus’s traditional habitat is disappearing. The walrus of the Pacific Arctic face an uncertain future. They might be able to move, they might be able to adapt. Or they might be stuffed. ■

➔ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xhunqv



A call to those who care about Europe's science

Better collaboration is a laudable goal, but that alone will not be enough to fix the damage caused by Europe's falling investment, says Amaya Moro-Martin.

When the European Parliament asked its proposed new commissioner for research what the continent should do about the state of its science, Carlos Moedas pledged greater cooperation between member states. Moedas might not have noticed, but we are already uniting: to protest against vicious budget cuts that are wrecking our scientific base and threatening our economic future.

These protests will reach a symbolic climax next week, with events planned in several European capitals, including the arrival in Paris of cycling French scientists involved in the *Sciences en Marche* campaign.

To mark this week of action and to highlight the need for a rethink on cuts, I and colleagues from across Europe have drafted an open letter to national governments and the European Parliament and Commission.

We encourage *Nature's* readers, as scientists and citizens who care about the future of research in Europe, to sign it here: openletter.euroscience.org.

The problems are many but can be summarized simply. The policy-makers and leaders of an increasing number of nations have completely lost touch with the reality of research.

They are ignoring how a strong research sector can contribute to the economy, something that is particularly crucial in the countries hit hardest by the economic crisis. Instead, they are imposing drastic budget cuts that are making these countries even more vulnerable. And all under the complacent gaze of European institutions.

There are too many examples to list, but here are some of the most prominent: since 2009, Italy has seen recruitment of scientists fall by 90% and the amount spent on basic research drop to nothing. In Spain, the amount of money spent on civilian research and development has dropped by 40%, and fewer than 10% of researchers who retire are being replaced. Since 2011, the budget of Greek research centres and universities has halved, with a freeze on hiring. Already reeling from budget cuts of 50% for universities and research centres, Portugal may now have to close half of its research units because of a flawed evaluation process supported by the European Science Foundation.

French researchers are alarmed by the 20–25% decline in the number of scientific and academic positions and by the less than 10% success rate of the increasingly more prevalent grant-based funding. Even Germany is fostering fixed-term contracts through its science-employment act, making the future uncertain even for very experienced researchers.

Most of these measures are in the name of austerity. Europe's vain hope is that the private sector will step up to provide the spending increases required to reach the Lisbon Treaty's goal of 3% of gross domestic product. But this ignores the fact that private backing tends to be spurred by public investment: more than half of the United States' economic growth has come

from innovation that has roots in federally-funded research.

The drastic budget and hiring cuts, the latter recommended by the European directive, are triggering a brain drain. Where they can, scientists are shifting from the less-affluent south to the north of Europe. Where they cannot, many are abandoning the continent altogether.

Spanish policy-makers are even denying the problem exists, despite the clear collapse in employment opportunities and the visible emptying of research centres. If the lack of opportunities continues, more and more European scientists will simply leave research.

Europe's research commissioner deserves a chance to improve the situation. But it is telling that Moedas's guidance letter from the commission's president-elect, Jean-Claude Juncker, does not mention the need

to address budget cuts or the brain drain. Instead, it asks him to focus on applied research, and in particular on boosting the participation of the private sector and of small and medium enterprises.

Despite what some politicians believe, applied research is unlikely to have much immediate impact on the market. Marketable research products are the low-hanging fruit of an intricate research tree, and undermining basic research will slowly kill the roots.

Ultimately, Europe's approach ignores how the scientific process works. Research requires experiments, and not all will be successful. Excellence is the tip of an iceberg: it is prominent only because of the support of the body of work beneath.

Instead, science funding at both the national and the continental levels is going to a diminishing number of well-established research groups.

This is not conducive to the diversified portfolio that Europe will need to face the societal and technological challenges of tomorrow. It also increases the gap between member states, because those well-funded research institutions are systematically recruiting a selected group of grant holders.

Research cannot follow political cycles: it is about investment in the future. And it should not just serve the economy, but also aspire to increase knowledge and social welfare, including for those with no resources to pay the bill.

Too many of those in positions of power in Europe have chosen to ignore this. We are determined to remind them. We call on you to help us. ■

Amaya Moro-Martin is an astrophysicist at the Space Telescope Science Institute (STSI) in Baltimore, Maryland, and a member of the governing board of Euroscience.

e-mail: amayamoramartin@gmail.com

The views and opinions expressed are those of the author and do not necessarily reflect the views of the STSI.

RESEARCH
CANNOT FOLLOW
POLITICAL
CYCLES: IT
IS ABOUT
INVESTMENT
IN THE
FUTURE.

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/remli

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

ELECTRONICS

Fluid-based sensor bends and twists

Electronic sensors made using liquids can outperform other flexible devices that have solid components.

Most sensors rely on solid metals that form junctions. To render such devices flexible, Ali Javey of the University of California, Berkeley, and his colleagues developed a way to make a junction between two different fluids that does not allow them to mix.

For the junction, the team fabricated a series of microchannels, each about 30 micrometres wide, which are designed to let in only one of the fluids — an ionic liquid. The other fluid, the commercial liquid metal Galinstan, has too much surface tension to enter the channels.

The sensors could detect humidity, oxygen and temperature; the temperature sensor was 17–46 times more sensitive than flexible alternatives made with solid components. The device could be useful in prostheses, robotics and smart wallpapers, the authors say. *Nature Commun.* 5, 5032 (2014)

GENOMICS

Hundreds of genes for height

One of the largest-ever genome-wide association studies has identified 697 genetic variants for human height — several hundred more than a previous, smaller study.

Peter Visscher of the University of Queensland in Brisbane, Australia, and a team of hundreds of scientists analysed the combined results

of 79 genome-wide association studies, encompassing 253,288 people, and found that the common gene variants account for 16% of the genetic contribution to height. Many of the variants lie near genes in biochemical pathways associated with skeletal growth, and others were linked to genes that were not previously thought to be growth-related.

The authors suggest that tens of thousands of common genetic variants together influence height, and that bigger genome-wide

association studies including hundreds of thousands of people will continue to provide useful biological information. *Nature Genet.* <http://doi.org/v6k> (2014)

CLIMATE CHANGE

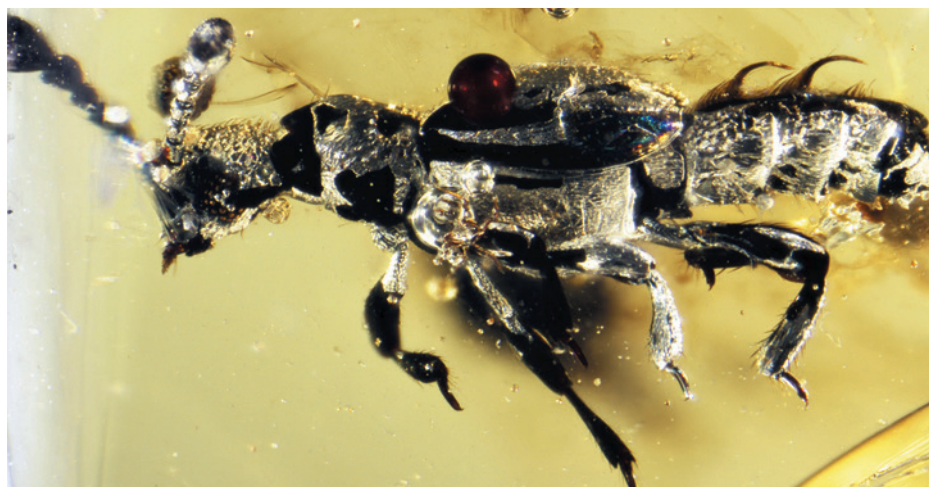
Ocean warming underestimated

Recent estimates of global temperature rises in the upper ocean may have been too low.

Oceans absorb the majority of the heat resulting from global warming. Paul Durack

of the Lawrence Livermore National Laboratory in California and his colleagues used a range of climate models and satellite observations to reassess observed changes in ocean warming between 1970 and 2004. They concluded that upper-ocean warming has been underestimated by as much as 58%, mainly because of sparse data from oceans in the Southern Hemisphere.

A separate study by William Llovel and his colleagues at NASA's Jet Propulsion Laboratory in Pasadena, California, showed that the



EVOLUTION

Oldest ant lover found entombed

Insects that embed themselves in ant colonies have existed for nearly as long as their hosts, and some have evolved rapidly, probably in response to ant diversification.

Joseph Parker at Columbia University and David Grimaldi from the American Museum of Natural History, both in New York, conclude that a 52-million-year-old beetle found entombed in amber (pictured) is the earliest known example of a myrmecophile — species that depend on ants for their survival. The newly identified species, *Protoclaviger trichodens*, is an extinct ancestor of contemporary rove beetles and has features

such as hairs to deliver tasty secretions to worker ants. It existed just as ant populations were starting to rise in tropical rainforests.

In another myrmecophile study, Wendy Moore and James Robertson at the University of Arizona in Tucson used DNA sequences to examine the evolutionary relationships of ant-nest beetles (*Paussus* spp). They found that the beetles are some of the fastest evolving animals on Earth. For example, the common ancestor of the 86 species that are native to Madagascar existed just 2.6 million years ago.

Curr. Biol. <http://doi.org/v6q>; <http://doi.org/v6r> (2014)

JOSEPH PARKER AND DAVID A. GRIMALDI

ocean's upper 2,000 metres have strongly warmed since 2005. However, at depths below 2,000 metres, the ocean has absorbed negligible amounts of heat during this period.

These two studies have implications for accurately assessing the effects of climate change on sea-level rise.

Nature Clim. Change <http://doi.org/v58>; <http://doi.org/v6j> (2014)

GEOPHYSICS

Mid-depth quakes are risky too

Earthquakes that originate at intermediate depths are an underappreciated seismic risk, according to a study of a June 2014 earthquake in the western Aleutian Islands off Alaska.

The epicentre of the magnitude-7.9 quake was approximately 100 kilometres deep, making the quake the largest in this depth range — between about 70 and 200 km down — in the past century. Thorne Lay of the University of California, Santa Cruz, and his colleagues analysed the earthquake and found that the energy release was weak at first but became strong during its final 25 seconds.

Other regions with a similar tectonic structure, such as Japan and Indonesia, should be aware of how big and powerful intermediate-depth quakes can be, the authors warn.

Geophys. Res. Lett. <http://doi.org/v45> (2014)

NEUROSCIENCE

How curiosity enhances learning

Curiosity boosts people's ability to learn and retain new information, thanks to key reward and memory centres in the brain.

Matthias Gruber and his colleagues at the University of California, Davis asked volunteers to rate their level of curiosity for a series of trivia questions, and then scanned

their brains as they saw the questions and waited for the answers.

For questions that they were curious about, participants remembered answers better than for questions in which they were less interested. Brain scans showed increased activity during this learning in regions that respond to reward and regulate memory formation, and revealed heightened connectivity between the two regions.

The volunteers were shown unrelated faces while they waited for the trivia answers, and were better at learning those faces when their curiosity was aroused. This suggests that curiosity also helps with the learning of incidental information.

Neuron <http://doi.org/v6m> (2014)

CLIMATE SCIENCE

Plant growth leads to Arctic warming

Increased carbon dioxide in the atmosphere is known to boost vegetation cover at high latitudes — and this could accelerate Arctic warming year-round.

Grasses and shrubs have a warming effect because plant-covered areas reflect less sunlight than barren surfaces do. Baek-Min Kim at the Korea Polar Research Institute in Incheon, South Korea, Sang-Yoon Jun at the Korea Institute of Atmospheric Prediction Systems in Seoul and their colleagues used a climate model to study the impact of doubled CO₂ concentrations and increased high-latitude plant growth on Arctic temperatures.

They found that increased vegetation in summer warms the surface and this heat moves to the Arctic, where it causes additional ocean warming and sea-ice melting in winter and spring. The exposed ocean then releases more heat, leading to a further boost in Arctic warming and

SOCIAL SELECTION

Popular articles
on social media

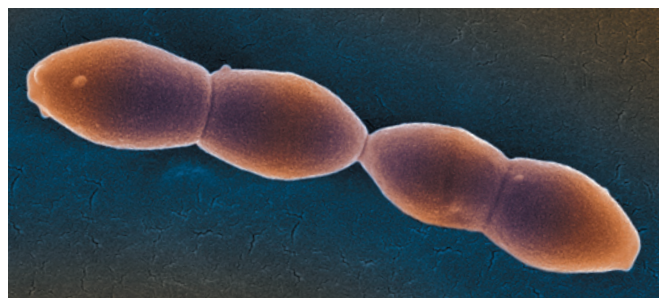
'Riff-raff' charge draws ire

When Steven McKnight, president of the American Society for Biochemistry and Molecular Biology (ASBMB), publicly complained in the society's September newsletter about the "riff-raff" that has infiltrated the research world (see go.nature.com/tu4nun), he quickly found himself in the social-media spotlight. He argued that the grant-review system run by the US National Institutes of Health is failing partly because of mediocre researchers serving on review committees. "The average scientist today is not of the quality of our predecessors," wrote McKnight, who chairs the biochemistry department at the University of Texas Southwestern Medical Center in Dallas.

Many researchers found his article insulting. Michael Hendricks, a neuroscientist at McGill University in Montreal,

Canada, tweeted: "If I had any idea what the @ASBMB was, I would be cancelling my membership today." In an interview, McKnight said that he regrets his choice of words, but stands by his arguments.

➔ NATURE.COM
For more on popular papers:
go.nature.com/e5exoh



promoting even more plant growth the following season, the team says.

Environ. Res. Lett. 9, 094007 (2014)

MICROBIAL GENETICS

Gene switch helps bacteria invade

A bacterium that causes pneumonia and other ailments can switch between six different forms by rearranging key genes, allowing the microbe to alter its ability to infect.

Streptococcus pneumoniae (pictured) lives harmlessly in the nose but can cause serious infections in some people. Michael Jennings at Griffith University in Southport, Australia, Marco Oggioni at the University of Leicester,

UK, and their co-workers focused on a specific set of genes comprising a system called SpnD39III in a strain of *S. pneumoniae*. They found that rearrangements of these genes result in six distinct bacterial subpopulations, each with its own pattern of methyl groups on DNA, which modify gene expression.

The subpopulations caused infections of varying severity in mice.

The finding suggests how this pathogen can quickly adapt to changing environments, such as when it shifts from harmless colonization to invasive disease.

Nature Commun. 5, 5055 (2014)

➔ NATURE.COM
For the latest research published by Nature visit:
www.nature.com/latestresearch

SEVEN DAYS

The news in brief

AWARDS

Nobel prizes

Three neuroscientists share this year's Nobel Prize in Physiology or Medicine for their insights into the neural basis of spatial navigation. John O'Keefe was honoured for his discovery of 'place cells', which are activated when a rat passes particular spots; May-Britt Moser and Edvard Moser discovered 'grid cells', which help to create a coordinate system in the brain. The physics prize was awarded to Isamu Akasaki, Hiroshi Amano and Shuji Nakamura for their invention of blue light-emitting diodes (LEDs), which enabled today's bright, energy-efficient white LEDs. See pages 152–154 for more. *Nature* went to press before the chemistry prize was awarded, but full details will be available at go.nature.com/ygtzbs.

EVENTS

Ebola exported

A health worker in Spain has tested positive for Ebola, the country's health minister said on 6 October. The worker, the first person thought to have contracted the virus outside Africa, had treated a missionary who had returned to Spain from Sierra Leone. On 30 September, health officials confirmed the first case of Ebola diagnosed in the United States — a man who arrived in Dallas, Texas, from Liberia on 20 September. On 6 October, the US Centers for Disease Control and Prevention said that the man may have exposed up to 48 people to the virus. See page 139 for more.

Animal care

A UK government report released on 2 October cleared Imperial College London of

nearly all allegations of animal cruelty raised by an animal-rights group in 2013. But the Home Office investigation noted persistent shortcomings in the management of animal care, and five instances of non-compliance — since addressed by the university — that were symptomatic of "a widespread poor culture of care" in the university's animal-research laboratories. See go.nature.com/wyc2vz for more.

Telescope turns on

A 12-metre radio telescope on Kitt Peak in Arizona has begun observations, scientists at the University of Arizona in Tucson said on 2 October. The telescope is one of three

prototypes originally made for the Atacama Large Millimeter/submillimeter Array (ALMA) in Chile, and will be used to study phenomena such as molecules in interstellar space and supermassive black holes. The university lost a bid for another of the prototype dishes in 2011; that antenna is destined for Greenland (see page 147).

FACILITIES

Nuclear-dump plan

The US Department of Energy issued a plan on 30 September to revive the country's only deep-storage repository for nuclear waste. Operations at the Waste Isolation Pilot Plant near Carlsbad, New

Mexico, were suspended after two incidents in February: an underground fire and the accidental release of some radioactive material (see *Nature* **509**, 267–268; 2014). The plan aims to resume at least some operations in 2016, and includes improving ventilation systems and retraining staff. The cost could top US\$500 million, according to the department.

Carbon capture

The world's first commercial coal-fired plant that can capture its carbon dioxide emissions was officially launched in Canada on 2 October. The refitted Boundary Dam Power Station in Saskatchewan — a Can\$1.3-billion



COREY ACCARDO/NOAA

Walrus seek shelter on land

A lack of sea ice has driven more than 35,000 Pacific walrus (*Odobenus rosmarus divergens*) onto land at Point Lay, Alaska, the US Geological Survey said on 1 October. The animals normally spend summers resting on ice floes, occasionally diving to the ocean floor to feed on clams, snails and worms. But with ice levels low in the Chukchi Sea this summer,

walrus have come ashore in record numbers (pictured). The animals are easily spooked, so scientists say that the risk of fatal stampedes is high. The US Fish and Wildlife Service is considering whether to protect the Pacific walrus under the Endangered Species Act owing to harm from hunting and sea-ice loss; a decision is expected in 2017. See page 140 for more.

(US\$1.2-billion) project — is designed to trap around 1 million tonnes of CO₂ a year. Most of the gas will be bought by an oil company, which will pump the compressed CO₂ deep underground to flush out stubborn oil reserves. See go.nature.com/nje9il for more.

FUNDING

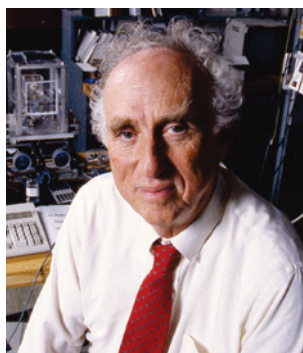
BRAIN bucks

The US National Institutes of Health has awarded its first US\$46 million in grants for the Brain Research through Advancing Innovative Neurotechnologies (BRAIN) Initiative. The agency announced 58 awards on 30 September, ranging in size from about \$370,000 to \$1.9 million. The White House also announced that two more government agencies — the US Food and Drug Administration and the Intelligence Advanced Research Projects Activity — would join the multi-agency effort. See go.nature.com/8trpex for more.

PEOPLE

Physicist dies

Martin Perl, who shared the 1995 Nobel Prize in Physics, died on 30 September, aged 87. Perl was honoured for his role in discovering the tau lepton, a subatomic particle that is 3,500 times



heavier than an electron. The discovery was the first sign of a previously unknown family of leptons, and filled in a missing piece of the standard model of particle physics. Perl (pictured) also received the Wolf Prize in Physics in 1982.

PhD pulled

The University of Constance in Germany was within its rights to revoke the PhD of physicist Jan Hendrik Schön, according to a final ruling announced on 1 October by a German constitutional court in Karlsruhe. In 2002, Schön was dismissed from Bell Laboratories in Murray Hill, New Jersey, for falsifying research findings, including several that were published in *Nature* and *Science*. Two years later, the university decided to revoke the doctorate that it had awarded to Schön in 1998 because of his actions at Bell Labs. See go.nature.com/kb89km for more.

POLICY

Stem-cell saga ends

Italy's health ministry will not support a trial of a controversial stem-cell therapy that it had promised last year — marking the end of a two-year battle between the therapy's inventor and Italian scientists who had declared the treatment ineffective and possibly dangerous. Health minister Beatrice Lorenzin announced her final decision on 2 October, on the basis of conclusions from an expert committee that was convened after a court ruled a previous committee had been illegally biased. See go.nature.com/zlryhz for more.

PNAS rules tighten

The *Proceedings of the National Academy of Sciences* (PNAS) closed one of its avenues to publication on 1 October. Editor-in-chief Inder Verma said that the journal will discontinue its 'pre-arranged editor' process, whereby authors could submit manuscripts through a member of the academy. See go.nature.com/szpjio for more.

US–India space pact

The US and Indian space agencies signed an agreement on 30 September to increase their collaborative efforts. A working group will coordinate observations between NASA's MAVEN mission and the Mars

COMING UP

15–16 OCTOBER

Ministers arrive in Pyeongchang, South Korea, to discuss progress towards world biodiversity targets, at the final stages of global talks between nations that have signed the Convention on Biological Diversity. go.nature.com/lvg1dl

Orbiter Mission of the Indian Space Research Organisation — which arrived at Mars on 21 and 24 September, respectively — and explore cooperative efforts on future Mars missions. They will also jointly launch an Earth-observing satellite in 2020.

Trial data freed

Clinical-trial data in the European Union are to be released to the public after the European Medicines Agency agreed a transparency policy on 2 October. As of 1 January 2015, the medical regulator will proactively publish clinical reports from companies seeking marketing authorization from the agency for medicines. The policy had been the subject of fierce debate, which pitted transparency campaigners against pharmaceutical companies. See go.nature.com/p7slov for more.

BUSINESS

Virology mega-deal

US pharmaceutical giant Johnson & Johnson said on 30 September that it will pay about US\$1.75 billion in cash to acquire Alios BioPharma, a private biotechnology company in South San Francisco, California. Alios is known for its work on treatments for viral infections, such as respiratory syncytial virus.

► NATURE.COM

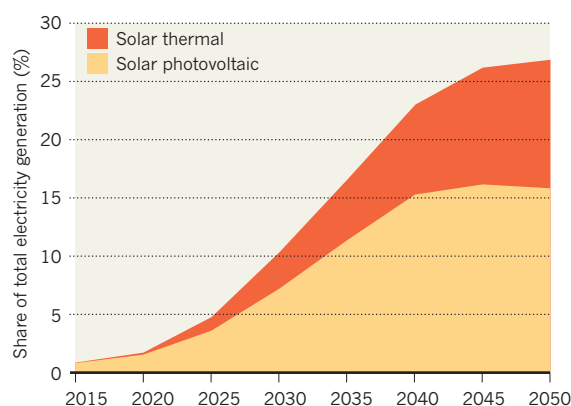
For daily news updates see:
www.nature.com/news

TREND WATCH

Solar power could provide 27% of the world's electricity by 2050, the International Energy Agency says — up from the 22% figure it suggested 4 years ago. In a strategy report published on 29 September, the Paris-based agency increased its projections for electricity supplied by photovoltaic modules from the 2010 value of 11% to 16%, because of the rapidly falling cost of the modules. The projections, however, assume the introduction of policies that encourage low-carbon electricity.

SOLAR SIGHTS SET HIGH

Photovoltaics and solar thermal energy could together become the world's leading source of electricity by 2050.



NEWS IN FOCUS

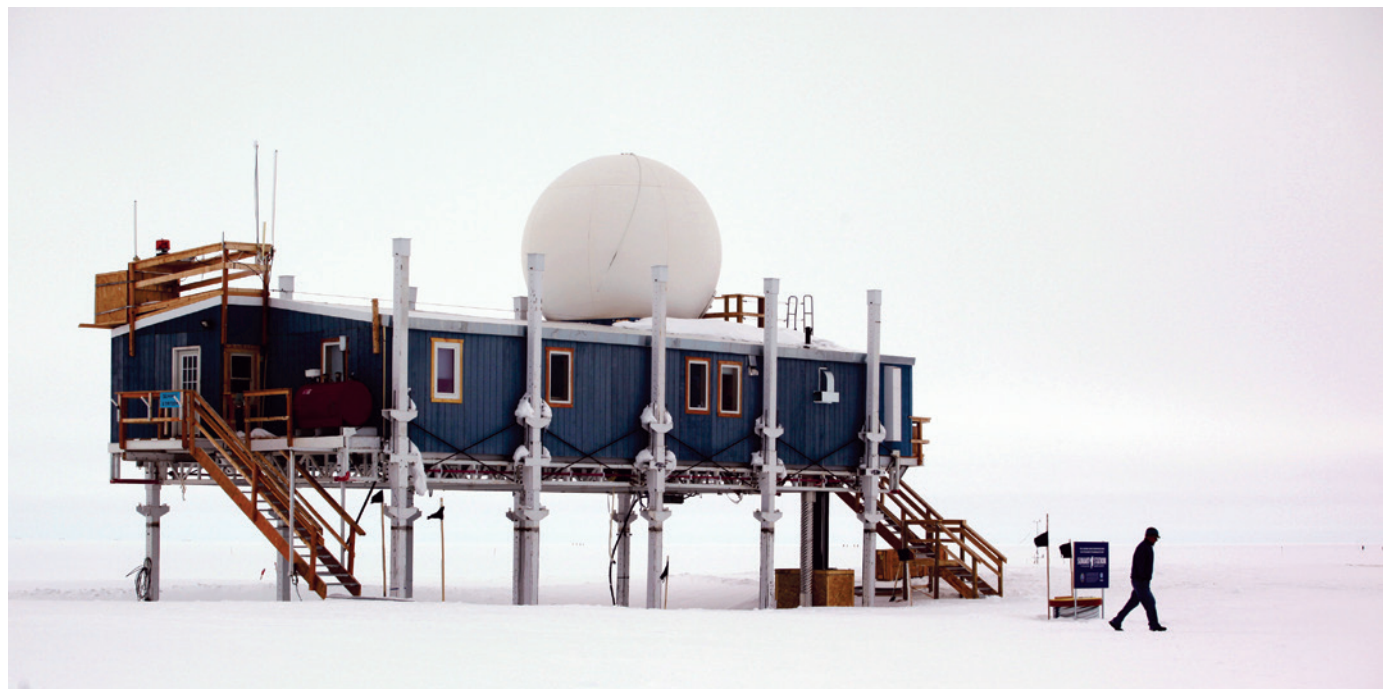
AUSTRALIA Scientists Down Under clash with politicians **p.148**

NEUROSCIENCE Japan's ambitious 10-year brain project begins **p.151**

NOBEL PRIZE Blue light-emitting diode scoops physics prize **p.152**



NEUROSCIENCE How the Mosers exposed the brain's sense of place **p.154**



BRENNAN LINSLEY/AP

Summit Station in Greenland, run by the US National Science Foundation, has experienced growing pains as the number of visiting researchers has increased.

FACILITIES

US plans upgrade for ageing Greenland research station

But proposals spur concern that development will pollute the nearly pristine site.

BY ALEXANDRA WITZE

Atmospheric researchers flock to Summit Station, on the peak of the Greenland ice sheet, for purity: at 72 degrees north and 3,216 metres above sea level, the air is about as pristine as can be. This is the premiere spot to grab an air sample uncontaminated by local emissions.

But more researchers are heading to the station these days, and with them come diesel generators, aeroplanes and snowmobiles. When the winds blow in the wrong direction, atmospheric scientists have to note in their data that a 'clean' sample might have been sullied by the very

machinery that made its collection possible.

And the problem is about to get worse. By 2017 or 2018, a US-Taiwanese collaboration plans to build a 12-metre radio telescope at Summit. The team says that it will take advantage of Greenland's phenomenal skies and allow them to measure a black hole's shadow with extraordinary precision. It will also require huge increases to staff numbers at the station and the site's power-generation capacity. Station planners must now work out how to balance the need for pristine environmental conditions against the emissions associated with the telescope itself and the station's rising population.

"This is a game-changer for Summit,"

says Brian Vassel, who manages atmospheric monitoring there and at five other global stations for the US National Oceanic and Atmospheric Administration's Earth System Research Laboratory in Boulder, Colorado. "How do we all play nice when we're at the top of the ice cap?"

Summit was established in 1989, with permission from the Danish and Greenlandic governments, to create a base camp for an ice-core drilling project. In the 1990s, the station began to host other research visitors, mostly atmospheric scientists and glaciologists. Since 2003, it has been occupied year-round.

But the facilities are less than ideal. The station consists of a cluster of poorly insulated ►

► buildings and laboratories, and berths so scarce that many scientists must sleep in tents at the peak of the summer crush, when some 45 researchers may be on site. In winter, at least five technicians must stay to maintain atmospheric measuring equipment. The winter staff will at least double when the radio telescope arrives, as will the station's power needs.

The US National Science Foundation (NSF), which pays for Summit's infrastructure and oversees much of its science, plans to adapt, in part, by spreading out the facilities. The agency has drafted a plan that would put the telescope more than a kilometre away from the current station, move the airstrip further east, and build an atmospheric observatory in the sector designated for clean air and snow measurements.

Greenland's excellent observing conditions make the development worth the trouble, says Paul Ho, an astronomer and former director of Taiwan's Academia Sinica Institute for Astronomy and Astrophysics (ASIAA). Operators also plan to link the telescope with the Atacama Large Millimeter/submillimeter Array (ALMA) in Chile, about 9,000 kilometres away, to produce combined observations much sharper than either facility could produce alone.

In fact, the huge Greenland radio dish was originally built as a prototype for ALMA. In 2011, a consortium led by the ASIAA and the Smithsonian Astrophysical Observatory in Cambridge, Massachusetts, obtained the telescope in a controversial deal that saw an instrument built with US funds transferred to an international consortium (see *Nature* 470, 14; 2011). Parts of the antenna are now in several countries as it gets adapted for polar conditions. Plans call for it to be shipped to Thule, Greenland, in the next few years and then hauled 1,200 kilometres to Summit.

ASIAA is trying to raise US\$10 million to \$15 million to help pay for the telescope's transport and setup, and a big part of the pitch includes a request for funding green-energy technologies. Ho says that astronomers are aware that the site needs to remain as clean as possible. "We want to preserve that, absolutely. In all ways possible," he says. Still, it is not clear what renewable source would work best. Solar energy cannot provide power during the long Arctic winter. A wind generator has been tested at the site, but the wind is gusty and unreliable.

The real obstacle to improving the station might be cost. The NSF is not releasing specific estimates, but it will probably have to squeeze its current Arctic facilities budget — around \$40 million annually — to find money for Summit improvements.

"If we can handle these conflicts and make sure that the emissions don't impact the atmospheric measurements, it will bring a higher profile to Greenland," says Jack Dibb, an atmospheric chemist at the University of New Hampshire in Durham and member of the NSF's scientific advisory group for Summit. "We can do it — if we spend the money." ■



A slashed budget sparked protests in June by government-funded scientists in Canberra.

POLICY

Australian cuts rile researchers

Political scorn on top of shrinking funds creates hostility between scientists and Tony Abbott's government.

BY DANIEL CRESSEY

Australia is booming. The country's economy is strong, and this year the Organisation for Economic Co-operation and Development put the nation at the top of its 'better life' index — an attempt to quantify well-being in industrialized countries.

But many scientists Down Under are not feeling on top of the world. Just over a year since taking office, the coalition government, under the leadership of the Liberal Party's Tony Abbott, has cut several basic-research programmes and put science under the auspices of industry minister Ian Macfarlane — who last month dismissed those who complained about the consolidation as "precious petals in the science fraternity". Last week, another government minister threatened that two key basic-research programmes could be targeted if parliament resists proposed cuts

in higher-education funding. This came just days after the Australian Greens party released an analysis produced by the Parliamentary Library showing that government spending on research and development fell to 0.56% of gross domestic product for 2014–15, its lowest level since 1989–90 (see 'Lean years').

"It is getting noticeably more difficult to get funding, with success rates for various fellowship and project grant schemes declining," says Darren Saunders, a medical researcher at the University of New South Wales in Sydney. "There is a very real perception that our current government doesn't value or respect science and scientists as highly as we would like."

The 2014–15 budget, released in May by the current government, cuts millions of dollars in research funding from government agencies such as the Commonwealth Scientific and Industrial Research Organisation (CSIRO), which employs thousands

CSIRO STAFF ASSOCIATION

of scientists at various locations. At CSIRO alone, hundreds of jobs will be lost this year and the organization's annual budget will be cut by Aus\$30.9 million (US\$27 million).

"I've been involved with CSIRO for 15 years and the feedback everyone is getting is that morale is at an all-time low," says Sam Popovski, a former agricultural scientist who is now a full-time trade-union representative for the CSIRO staff association.

Popovski points out that the previous government also made cuts to the CSIRO, which by mid-2015 will have lost one in five staff over the past two years, he says. Of roughly 1,300 job losses, Popovski estimates that more than half can be attributed to the Abbott administration, with another 300 caused by cuts from the previous government and 300 as a result of an internal reorganization.

DOWN AND OUT

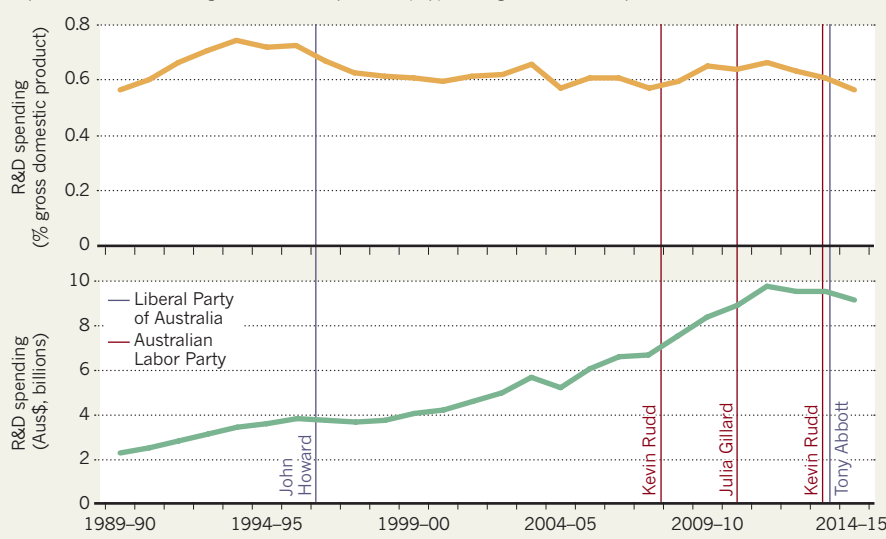
The government has also cut the budget of the Australian Research Council, the main source of non-medical government research grants. Leslie Field, secretary for science policy at the Australian Academy of Science and vice-president of the University of New South Wales, says that the move is likely to reduce grant success rates below their current level of around 20%, thus eliminating support for a significant share of Australia's high-quality research. On the government's general science spend, Field says: "There is no question that the cuts will have a negative effect on Australia's science system."

Proposed changes to the nation's higher-education system threaten to make a difficult funding situation even tougher. The government is seeking to cut hundreds of millions of dollars from funding for student places, and wants to remove a cap on student fees to allow universities to claw back the difference. But the proposal has run into trouble in parliament. In response, education minister Christopher Pyne said in parliament on 2 October that if the reform does not pass, the government could reduce scientific programmes to compensate. Cuts to the National Collaborative Research Infrastructure Strategy, which supports scientific facilities, and the Future Fellowships programme, designed to support 'outstanding' mid-career researchers, would cost up to 1,500 jobs, he said.

Because universities can use student fees to subsidize their research, doubt over student

LEAN YEARS

The Australian government's spending on research and development (R&D) has stagnated in real terms since 2011–12 (bottom), and has fallen to its lowest level in 25 years when expressed as a share of gross domestic product (top). Changes in leadership are also shown.



funding further threatens the scientific community, says Brian Schmidt, an astronomer at the Australian National University in Weston Creek and a Nobel laureate. "There's huge uncertainty right now," he says — especially for early-career researchers.

Not all areas of science have faced cuts in the past year. The Abbott government has strongly backed medical research, announcing that it wants to create a billion-dollar Medical Research Future

"There is no question that the cuts will have a negative effect on Australia's science system."

Fund, which would invest Aus\$20 million for 2015–16, rising to Aus\$1 billion by 2022. Although this potential extra funding — which must still be approved by parliament — is welcomed by medical researchers, there is concern over the proposal that this money would be raised by a Aus\$7 fee for people to visit a doctor.

"That medical-research fund of course is welcomed by the medical-research community," says Peter Doherty, an immunologist at the University of Melbourne and a Nobel laureate. But he adds that "the funding mechanism has created a lot of hostility" towards medical research. Charities have expressed

concern that public donations could decline as a result of the charge.

Some researchers say it is telling that the Abbott government has done away with a dedicated science minister, instead making industry minister Macfarlane responsible. Last month, Macfarlane lashed out at those who complained about his dual role with his now-famous 'precious petals' quote. He took umbrage at what he said was the insinuation that he did not care enough about science by saying: "I'm just not going to accept that crap." In a e-mail to *Nature*, Macfarlane said: "It's no accident that science is in the industry portfolio." He added that the government would soon release a 'competitiveness agenda', of which a key element would be creating stronger links between business and research.

Schmidt calls the 'precious petals' comment "poorly judged", but puts it down to ongoing "name-calling" between the government and its critics in the science community.

Other researchers are angrier. The comment inspired US researcher Ainsley Seago, who studies beetles at Macquarie University in Sydney, to create 'I'm a precious petal' T-shirts for irate scientists. Macfarlane's quote "was terrifically insulting", she says, and reminiscent of the "science-hostile Bush administration". ■



EXPLAINER

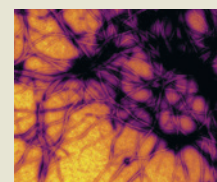


How disease detectives are working to contain Ebola in West Africa
go.nature.com/5sofin

MORE NEWS

- Gravity map uncovers sea-floor surprises go.nature.com/oyfyio
- Moon's largest plain is not an impact crater go.nature.com/tv3f2n
- Disclosing conflicts of interest has unintended effects go.nature.com/if5llc

DISEASE



Computer scientist makes prion advance
go.nature.com/1mhzc



A mock-up of a torsion balance used by British natural philosopher Henry Cavendish to measure G in 1798.

PHYSICS

Rivals join forces to nail down Big G

Metrologists meet to design the ultimate gravitational-constant experiment.

BY ELIZABETH GIBNEY

It is one of nature's most fundamental numbers, but humanity still doesn't have an accurate value for the gravitational constant. And, bafflingly, scientists' ability to pinpoint G seems to be getting worse. This week, the world's leading gravity metrologists are meeting to devise a set of experiments that will try to set the record straight. This will call for precision measurements that are notoriously difficult to make — but it will also require former rivals to work together.

In his 1687 *Philosophiæ Naturalis Principia Mathematica*, Isaac Newton outlined his theory of gravity: that the force between two objects is proportional to the product of their masses and inversely proportional to the square of the distance between them. G is the constant by which the masses must be multiplied to put an absolute value on that attraction. But more than 300 years later, the constant — known as Big G to distinguish it from little g , the acceleration due to gravity on

Earth — is known for sure to only 3 significant figures ($6.67384(\pm 0.0008) \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2}$).

Independent groups of physicists have been trying to pin down the true value of the constant for decades, but in recent years, rather than converging on an ever more precise figure, the results of their experiments have been diverging, causing the uncertainty in the official figure to rise (see 'Trouble with Big G '). "There's no other fundamental constant of physics where we've had such a wide dispersion of results," says Terry Quinn, former director of the International Bureau of Weights and Measures (BIPM) in Paris. Some or all of the experiments must either grossly underestimate their uncertainty or miss significant systematic errors, he adds.

On 9 and 10 October, metrologists will meet at the US National Institute of Standards and Technology (NIST) in Gaithersburg, Maryland, to devise a plan to resolve the issue.

Gravity is hard to measure because it is an extremely weak force. In a tussle over a paper clip, for example, the electromagnetic force of a small magnet beats the gravitational

pull of the whole of planet Earth. To measure G , scientists must eliminate every other potentially perturbing force. The oldest, and still most common, method measures the tiny rotation of a torsion balance — a rod hanging on a wire — caused by the attraction between masses at either end of the rod. Others include measuring the movement of pendulums — or the change in weight of test objects suspended on a 'balance beam' — in response to the presence of giant masses, and measuring the acceleration of cold rubidium atoms.

A big item on the agenda at the meeting will be debating how to choose a small number of these experiments to be conducted by members of a consortium, this time with an unprecedented level of oversight. Each experiment will be repeated by two independent groups, using identical sets of equipment created and tested at a third institution. While the experiments are going on — and there is still time to fix them — experts from outside those two groups will hunt for errors. In the past, says Quinn, who is a driving force behind the NIST meeting, scientists have picked holes in each other's experiments only after they were published, making it difficult to verify whether those problems were really the cause of an error.

Although meeting attendees are enthusiastic about the project, picking the experiments will not be easy, because each scientist is likely to gun for their own preferred method. "I've been working for 12 years on my experiment," says Andrea De Marchi, a physicist at the Polytechnic University of Turin in Italy, who will be presenting his technique to the meeting. "I guess everybody else has too."

Until now, scientists measuring G have competed; everyone necessarily believes in their own value, says Stephan Schlamminger, an experimental physicist at NIST. "A lot of these people have pretty big egos, so it may be difficult," he says. "I think when people agree which experiment to do, everyone wants their idea put forward. But in the end it will be a compromise, and we are all adults so we can probably agree."

Working together could even be a stress reliever, says Jens Gundlach, an experimental physicist at the University of Washington in Seattle. Getting a result that differs from the literature is very uncomfortable, he says. "You think day and night, 'Did I do everything right?'"

Cash is another hurdle. Big G experiments have often piggybacked on other projects. Justifying funding for them can be tough, because there are currently no theories or other experiments that rely on G having a particular value: in determining the orbits of satellites or planets, for example, only the result of multiplying the gravitational constant by the mass of the central body is relevant, and that can be determined precisely enough without knowing G .

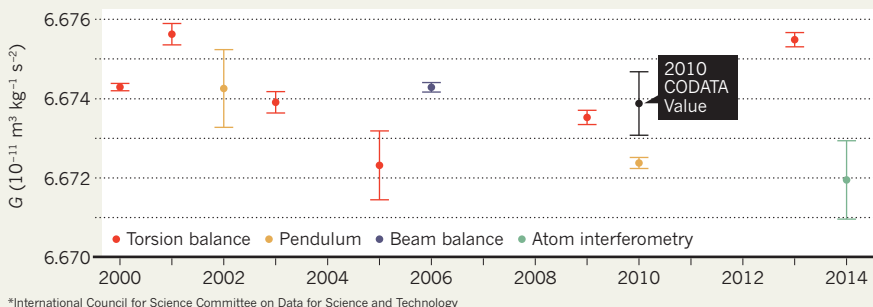
Big G might yet find a use in quantum gravity, the attempt to reconcile the theory of gravity with quantum mechanics, but for now,

the main motivation for chasing the constant is simply that not having a reliable value is irksome. “It’s hard to live with discrepancy,” says Schlamminger. It is possible that G appears to be so capricious because our basic understanding of gravity is flawed, he says, but that seems unlikely.

Quinn points out that NIST is hosting this week’s meeting, and he hopes that the institute will provide initial funding for the project, but nothing is yet firm. Assuming that cash is found, the consortium could have a reliable value for G in the next five years, says Quinn, or at least a good handle on where disagreements lie. “There must be a solution to this,” he says, “there must be.” ■

TROUBLE WITH BIG G

In 2000, scientists measured the gravitational constant, G , with smaller error bars than ever before. But since then, a variety of experiments using different techniques have produced a range of values — and uncertainty in the official CODATA* value has increased since 2006.



NEUROSCIENCE

Marmosets are stars of Japan's ambitious brain project

Ten-year brain-mapping effort will use monkeys to study human neural and mental disorders.

BY DAVID CYRANOSKI

Europe has one, the United States has one. Now Japan has thrown its hat into the neuroscience ring with the launch of its own brain-mapping project.

Unlike its Western counterparts, Japan's effort will be based on a rare resource — a large population of marmosets that its scientists have developed over the past decade — and on new genetic techniques that might be used to modify these highly social animals. The goal of the ten-year Brain/MINDS (Brain

Mapping by Integrated Neurotechnologies for Disease Studies) project is to map the primate brain to accelerate understanding of human disorders such as Alzheimer's disease and schizophrenia. On 11 September, the Japanese science ministry announced the names of the group leaders — and how the project would be organized.

Funded at ¥3 billion (US\$27 million) for the first year, probably rising to about ¥4 billion for the second, Brain/MINDS is a fraction of the size of the European Union's Human Brain Project and the United States' BRAIN (Brain

Research through Advancing Innovative Neurotechnologies) Initiative, both of which are projected to receive at least US\$1 billion over the next decade. But researchers involved in those efforts say that Brain/MINDS fills a crucial gap between disease models in smaller animals that too often fail to mimic human brain disorders, and models of the human brain that need validating data.

“It is essential that we have a genetic primate model to study cognition and cognitive brain disorders such as schizophrenia and depression, for which we do not have good mouse models,” says neuroscientist Terry Sejnowski at the Salk Institute in La Jolla, California, who is a member of the National Institutes of Health BRAIN Initiative Working Group. “Other groups in the United States and China have started transgenic-primate projects, but none is as large or as well organized as the Japanese effort.”

Neuroscientist Henry Markram at the Swiss Federal Institute of Technology in Lausanne, who heads the European initiative, also welcomes the effort: “It is absolutely impressive, and Japan should be applauded for putting together an incredible plan.”

Central to the Brain/MINDS effort is the creation of transgenic marmosets to elucidate cognitive function and as models of human brain disease. Although more distantly related to humans than primates such as chimpanzees, these monkeys are in many ways ideal for brain research. Their small size and fecundity make them easier and more efficient to work with



Marmosets share behaviours, such as making eye contact as a means of communication, with humans.

than, say, macaques, which are commonly used as animal models.

The marmoset brain is compact, too — just 8 grams, making it relatively easy to analyse. Yet the frontal lobe — a brain area involved in human psychiatric disease — is more developed in marmosets and more like the human version than the frontal lobes of other animals with similarly small brains. Marmosets also share behavioural characteristics with humans that other monkeys, and even chimps, do not, such as living in family units that resemble human ones and making eye contact as a means of communication, rather than as a form of aggression.

The monkeys are thus expected to be a good model for studying conditions such as Parkinson's disease and Alzheimer's disease. And learning what causes the breakdown of social behaviours such as making eye contact might help to clarify the mechanisms underlying autism, suggests Katsuki Nakamura, a neuroscientist at Kyoto University in Japan who works with marmosets.

Recent advances could also help researchers to genetically engineer marmosets more efficiently. In 2009, a team led by transgenic-animal specialist Erika Sasaki at the Central Institute for Experimental Animals in Kawasaki, Japan, was the first to introduce a gene into a primate that was passed on to viable offspring.

A gene-editing method known as CRISPR now permits precise DNA alterations^{1,2} and has been used to create transgenic monkeys³.

At the same time, geneticists have identified some of the key mutations that contribute to human disorders such as schizophrenia and autism. "This is a very important time to be developing primate genetic models," says Robert Desimone, director of the McGovern

"I would love to see one single important disease studied in great depth."

Institute for Brain Research at the Massachusetts Institute of Technology in Cambridge.

Brain/MINDS will be divided into three groups. One, led by project co-leader Hideyuki Okano at Keio University, will map brain function to structure using techniques such as functional magnetic resonance imaging. The group will use transgenic models of disease to link studies of macro-scale functions — long-distance pathways across the brain — to micro-scale studies that identify and characterize specific neuronal features. The second group, led by the project's other co-leader, Atsushi Miyawaki at the RIKEN Brain Science Institute in Wako, and consisting of 17 independent teams, will develop technologies to support the mapping effort. And the University of Tokyo's

Kiyoto Kasai will lead the third group, which will collect information, such as brain scans, from human patients. This will can be used to find signatures of psychiatric, neurological and vascular disease to feed back to the marmoset researchers.

Japan's rules regarding the use of primates in such research are less stringent than European or US regulations, but the project could still face ethical hurdles over issues of cruelty to animals. "It will be important for the Japanese marmoset project to carefully examine the ethical issues that will inevitably arise up the road," says Sejnowski.

BRAIN/minds is technically ambitious. Neuroscientist Afonso Silva at the National Institute of Neurological Disorders and Stroke in Bethesda, Maryland, who collaborates with Sasaki, points out that CRISPR has yet to be shown to work in marmosets. He adds that the project's researchers must resist the temptation to tackle too many conditions, so as not to spread themselves too thinly. "I would love to see one single important disease studied in great depth," he says. "Validating the approach on that one disease will then enable the project to be repeated in other diseases." ■

1. Cong, L. *et al. Science* **339**, 819–823 (2013).
2. Mali, P. *et al. Science* **339**, 823–826 (2013).
3. Niu, Y. *et al. Cell* **156**, 836–843 (2014).

NOBEL PRIZE

Blue LED wins physics Nobel

Invention revolutionized lighting and will reduce global electricity consumption.

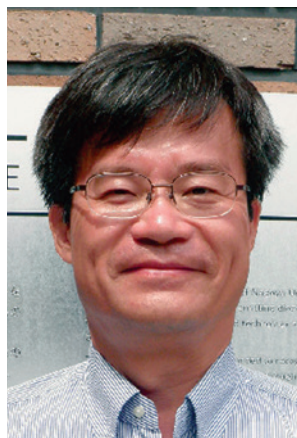
BY ELIZABETH GIBNEY

Found in smartphones, computer screens and energy-efficient bulbs, blue light-emitting diodes (LEDs) are everywhere. That they have now earned three Japanese-born inventors the 2014 Nobel Prize in Physics is a rare example of the award being given for a practical invention.

LEDs are devices that emit light when subjected to an electric current. Green and red versions have been around since the 1950s, but creating blue-emitters was a technical challenge that stumped industry efforts for decades.

It was not until the 1990s that engineers Isamu Akasaki and Hiroshi Amano, both at Japan's Nagoya University, in parallel with electrical engineer Shuji Nakamura, then working at Japanese chemicals firm Nichia, succeeded in creating a blue LED.

LEDs are sandwiches of semiconductor materials. The layers are 'doped' with other elements, which provides some layers with extra electrons and others with a surplus of 'holes',



Shuji Nakamura, Hiroshi Amano and Isamu Akasaki (left to right) won the 2014 Nobel Prize in Physics.

where missing electrons leave behind a positive charge. When an electrical current is applied, the electrons and holes combine at the junctions between the layers and emit light as a result.

From the 1980s, physicists focused their efforts on the material gallium nitride as a

target for making high-power blue LEDs, but they were faced with several technological hurdles. One was creating thin, high-quality crystals of the material, which are notoriously difficult to grow. Another was doping gallium nitride such that it emitted light efficiently.

JU PRESS/AFP/GETTY

Akasaki, Amano and Nakamura persisted with gallium nitride long after their competitors had moved on to other materials, says Wolfgang Schnick, a materials chemist at the Ludwig Maximilian University of Munich in Germany.

Their success in overcoming those hurdles has opened the door to white LEDs, which can have efficiencies nearly 20 times those of conventional bulbs.

Almost all white LED-based lights consist of a blue LED chip combined with one or more luminescent materials, which convert part of the blue light to longer wavelengths. "This has led to a revolution in the lighting industry, and will have more and more impact on the way people are lighting their homes," says Dirk Poelman, a materials scientist at Ghent University in Belgium.

Schnick says that the development "cannot be estimated too highly". "This will help to save up to 20% of the global electricity consumption," he says.

Schnick added that in future, blue LEDs are likely to find uses in portable devices that can disinfect or sterilize water, and perhaps in computer memories that use light instead of electricity to store data. Blue lasers — also invented by Akasaki and Amano, and separately by Nakamura — are already used in Blu-ray Disc technology.

The story is not without twists. Nakamura, who left Japan in 2000 to join the University of California, Santa Barbara, sued Nichia in 2001 over the scant compensation he received for inventing the blue LED technology while he worked there. The case was settled in January 2005, when Nakamura accepted ¥840 million (US\$8.1 million at the time). "Nakamura was quite determined to show that gallium nitride could be an effective LED technology. He pushed it very hard — it was something they were not initially concentrating on," says Martin Dawson, a photonics researcher at the University of Strathclyde in Glasgow, UK.

Speaking on a crackling telephone line to journalists in Sweden on 7 October, Nakamura said that the feeling of winning the prize was "unbelievable". Staffan Normark, permanent secretary of the Royal Swedish Academy of Sciences, told journalists that the trio had not been expecting the prize. "They had not been waiting all day and all night for this call," he said.

Announcing the prize, Per Delsing, chairman of the academy's Nobel Committee for Physics, said that the award cherished the tradition of its founder, the engineer and inventor Alfred Nobel. "I really think that Alfred Nobel would have been happy about this prize," he said. ■

Additional reporting by Richard Van Noorden.

NOBEL PRIZES

Prize for place cells

Discoverers of brain's navigation system get physiology Nobel.

BY ALISON ABBOTT & EWEN CALLAWAY

Brain cells that make up the biological equivalent of a satellite-navigation system have garnered three scientists the 2014 Nobel Prize in Physiology or Medicine. The discovery of the cells sheds light on one of neuroscience's great mysteries — how we know where we are in space.

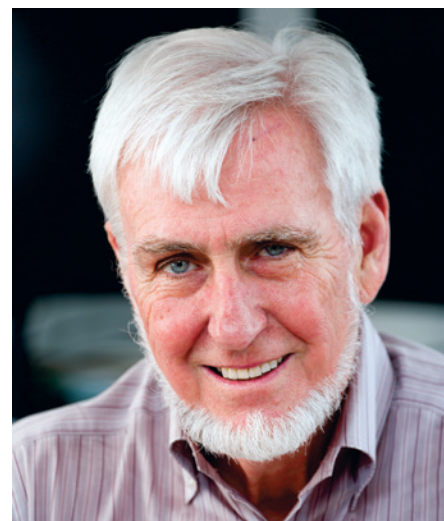
John O'Keefe of University College London won half of the prize for his discovery in 1971 of 'place' cells in the hippocampus, a part of the brain associated with memory. Edvard and May-Britt Moser, who are married and jointly run a lab at the Kavli Institute for Systems Neuroscience in Trondheim, Norway, share the other half for their 2005 discovery of 'grid' cells in an adjacent brain structure, the entorhinal cortex. Along with other navigation cells, grid and place cells allow animals to keep track of their position. Both cell types were discovered in rats, but have since been found in humans.

"Understanding where we are in space is one of the most fundamental issues for survival," says Tobias Bonhoeffer, director of the Max Planck Institute of Neurobiology in Martinsried, Germany.

The discoveries will also be key to answering the broader question of how the brain makes sense of the world, says neuroscientist Botond Roska of the Friedrich Miescher Institute for Biomedical Research in Basel, Switzerland. "These are three deep-thinking people who have changed the way we think about the brain," he says.

Most neuroscientists once doubted that brain activity could be linked with behaviour, but in the late 1960s, O'Keefe began to record signals from individual neurons in the brains of rats moving freely in a box. He put electrodes in the hippocampus and was surprised to find that individual cells fired when the rats moved to particular spots. He concluded that the memory of an environment may be stored as a specific combination of place-cell activities in the hippocampus (J. O'Keefe and J. Dostrovsky *Brain Res.* **34**, 171–175; 1971). "I realized that if you put them all together, you could have something like a map," says O'Keefe.

Fast-forward to the 1990s, and his work attracted the attention of the Mosers, then PhD students at the University of Oslo. They joined him in London as postdocs, but within months they had moved to the Norwegian University of Science and Technology in Trondheim to set up their own lab. There they discovered that some cells in the entorhinal cortex fire



John O'Keefe, co-winner of a 2014 Nobel prize.

when rats pass the points of a hexagonal grid. They found out that the brain uses this pattern as a coordinate system for spatial navigation (T. Hafting *et al. Nature* **436**, 801–806; 2005).

The pattern constitutes what is known as a neural code. It is the only one known to be generated entirely in the brain, marking a milestone for computational neuroscience (see page 154).

Both place and grid cells have practical relevance. The early stages of Alzheimer's disease affect the entorhinal cortex, and one of the first symptoms is losing one's way. The disease goes on to devastate the hippocampus, stripping sufferers of their memories. "It is a good example of how very basic research can help us gain the deeper understanding we need in such devastating diseases to move towards therapies," says Richard Morris, a memory researcher at the University of Edinburgh, UK.

May-Britt was presiding over a lab meeting when the call came from the Nobel committee in Stockholm. "I hesitated to answer it," she told *Nature*, laughing. "But I did — and I couldn't believe it; I even cried." Edvard's excitement was delayed: he was on an aeroplane to Munich, Germany, when his wife got the call. O'Keefe heard the news while working on a grant revision at home. "I'm totally delighted and thrilled," he said in front of a phalanx of television cameras at a London press conference.

The Mosers once described their time in O'Keefe's lab as "probably the most intense learning experience in our lives". O'Keefe has similar memories. "It was intense — because they're intense. They're absolutely superb scientists." ■

BRAINS OF NORWAY

NOBEL PRIZEWINNERS MAY-BRITT MOSER AND EDVARD MOSER HAVE SPENT A CAREER TOGETHER NEAR THE ARCTIC CIRCLE EXPLORING HOW OUR BRAINS KNOW WHERE WE ARE.

BY ALISON ABBOTT



The fact that Edvard and May-Britt Moser have collaborated for 30 years — and been married for 28 — has done nothing to dull their passion for the brain. They talk about it at breakfast. They discuss its finer points at their morning lab meeting. And at a local restaurant on a recent summer evening, they are still deep into a back-and-forth about how their own brains know where they are and will guide them home. “Just to walk there, we have to understand where we are now, where we want to go, when to turn and when to stop,” says May-Britt. “It’s incredible that we are not permanently lost.”

If anyone knows how we navigate home, it is the Mosers. They shot to fame in 2005 with their discovery of grid cells deep in the brains of rats. These intriguing cells, which are also present in humans, work much like the Global Positioning System, allowing animals to understand their location. The Mosers have since carved out a niche studying how grid cells interact with other specialized neurons to form what may be a complete navigation system that tells animals where they are going and where they have been. Studies of grid cells could help to explain how memories are formed, and why recalling events so often involves re-envisioning a place, such as a room, street or landscape.

While pursuing their studies, the two scientists have become a phenomenon. Tall and good-looking, they operate like a single brain in two athletic bodies in their generously funded lab in Trondheim,

Norway — a remote corner of northern Europe just 350 kilometres south of the Arctic Circle. They publish together and receive prizes as a single unit — most recently, the Nobel Prize in Physiology or Medicine, which they won this week with their former supervisor, neuroscientist John O’Keefe at University College London. In 2007, while still only in their mid-40s, they won a competition by the Kavli Foundation of Oxnard, California, to build and direct one of only 17 Kavli Institutes around the world. The Mosers are now minor celebrities in their home country, and their institute has become a magnet for other big thinkers in neuroscience. “It is definitely intellectually stimulating to be around them,” says neurobiologist Nachum Ulanovsky from the Weizmann Institute of Science in Rehovot, Israel, who visited the Trondheim institute for the first time in September.

The Mosers’ work has also given them traction at one of the most challenging twenty-first-century research frontiers: how the brain computes. Just as computers use programming languages such as Java, the brain seems to have its own operating languages — a bewildering set of codes hidden in the rates and timing with which neurons fire as well as the rhythmic electrical activities that oscillate through brain circuits. These codes allow the brain to represent features of the external world — such as sound, light, smell and position in space — in a language that it can understand and compute. With their grid-cell work, the Mosers have been the first to crack one such code deep in the

GEIR MOGEN/NTNU



brain; now the challenge for the field is to find all the rest.

“May-Britt and Edvard’s research lies at the very heart of the cognitive-neuroscience enterprise,” says Stanislas Dehaene, who studies consciousness at the Collège de France in Paris. “They are trying to understand the neural codes for cognition — and so unite biology with computer science and even philosophy.”

STARS ALIGN

The Mosers grew up on different Norwegian islands in the North Atlantic, where summer days seem eternal and the long winter nights are brightened only by the dancing Northern Lights. They were both from non-academic families and they went to the same school. But they didn’t get to know each other until 1983, when both were at the University of Oslo, both were wondering what to study and both were starting to realize that their true passion was for neuroscience and the brain.

Suddenly, everything sparked: romance between the two of them, intellectual curiosity and the beginnings of their mission in life — to find out how the brain generates behaviour. The Mosers visited one of the university’s more famous faculty members, electrophysiologist Per Andersen, and asked to do their undergraduate projects with him. Andersen was studying the activity of neurons in the hippocampus — a brain area associated with memory — and the two students wanted to try to link this precise activity of cells with

the behaviour of animals. Andersen, like most neuroscientists at the time, was sceptical about making such a big leap across the black box of the brain. But the pair wouldn’t leave his office until he gave in and offered them an apparently simple project: how much of the hippocampus could you cut away before a rat could no longer remember new environments?

The two young scientists embraced the challenge, and soon discovered something profound. Until then, it had been assumed that the hippocampus was homogeneous. But the Mosers showed that one side of it was much more important for spatial memory than the other side¹. That brought home to them the importance of detailed brain anatomy for understanding brain function, a lesson that would prove invaluable later in their careers.

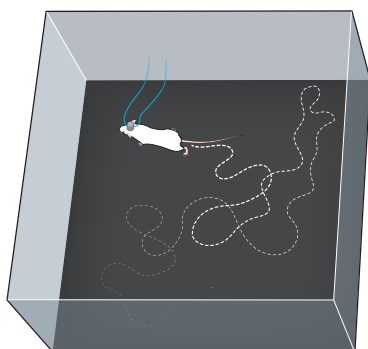
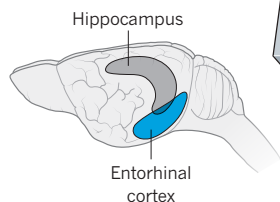
In 1984, while still undergraduates, the couple got engaged on top of the dormant volcano Mount Kilimanjaro in Tanzania. (The bitter temperature at the peak forced them to rush their exchange of rings, the quicker to get their gloves back on.) The pair had decided how their joint lives should be: children early, postdoc experience abroad and then their own lab together, somewhere in the world. These plans panned out — just a little faster than they had anticipated. Even before defending their PhDs, they accepted side-by-side postdocs in O’Keefe’s lab in London.

In the 1970s, O’Keefe had discovered neurons called place cells in

May-Britt Moser, Edvard Moser and their rats are exploring the brain’s navigation system.

A SENSE OF PLACE

Edvard and May-Britt Moser study grid cells in the brain's entorhinal cortex that help animals to understand where they are.



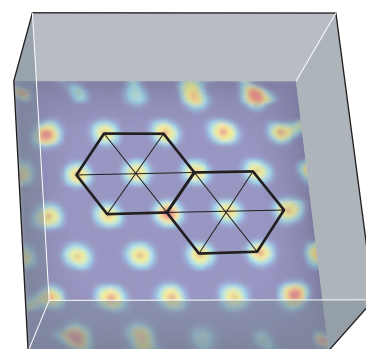
RAT ON THE RUN

The Mosers insert electrodes into a rat's entorhinal cortex and measure electrical signals from individual grid cells as the rat runs around a box, eating chocolate treats.



FIRING PATTERN

A single grid cell fires when a rat crosses certain points on the floor; it turns out that these points form a hexagonal grid, like a honeycomb.



POSITIONING SYSTEM

A hexagonal pattern gives the highest possible spatial resolution with the fewest cells. Each cell generates its own grid, and these overlapped patterns help the animal to recognize its location and direction.

the hippocampi of rats. These cells fire only when an animal is in a particular place — close to an exercise wheel, for example, or in front of a door. (Since then, other navigation-related neurons have been discovered, including those that fire when the head turns in a particular direction, or when a border, such as the long edge of a cage, is in view.) The research area was red hot, and the Mosers wanted to extend it.

But in 1996, just a few months into their postdocs, the Mosers received a surprise offer of two associate professorships at the Norwegian University of Science and Technology in Trondheim. They weren't sure about accepting: it would mean striking out alone, in a small university in a country isolated from the world's major centres of research. "But the offer of two posts in the same place and in the same research area was too good to turn down," Edvard says. They flew back home, by this time trailing a toddler and baby.

It wasn't easy to get established in Trondheim. They had to build a lab from scratch in a small basement, and establish an animal facility too. But only a few years in, they were winning big grants from the European Commission and the Research Council of Norway. And by then, the results were coming through.

ON THE GRID

The pair's first aim in Trondheim had been to better describe the origin of the place-cell signal. Although the cells themselves were in the hippocampus, it could be that cells elsewhere were instructing them when to fire. Remembering their lesson from the undergraduate lab, the Mosers knew that they needed to understand the brain's anatomy to see how the signals flowed physically across it.

In the lab, they adapted the standard experimental technique for studying place cells: implanting electrodes directly into a rat's hippocampus and recording from them as the animal runs freely in a large box (see 'A sense of place'). The electrodes — which are sensitive enough to pick up activity from single neurons — feed into a computer and map the exact spot on the floor of the box where each neuron fires. This appears on screen as a black dot. To make sure that the rat covers the entire floor area, the researchers scatter chocolate treats across it. (May-Britt is a chocolate enthusiast both inside and outside the lab.)

The Mosers chemically inactivated different parts of the hippocampus and its surroundings in the rat brains, and then tested whether the place cells continued to fire normally. In this way, they discovered² that information flowed to the place cells from the entorhinal cortex, a narrow strip of tissue running vertically up the lower back of the rat's brain. No one had paid much attention to this structure before, in large part because it is extremely difficult to access. One side lies very close to a large blood chamber; puncturing that would be fatal. The Mosers consulted an expert in neuro-anatomy and concluded that, fortunately, the ideal place for the

electrodes would be away from the chamber and close to the brain's surface. Then they started to repeat their experiments, recording from single neurons in the entorhinal cortex. That is when they found something unexpected.

The researchers saw that some of these entorhinal neurons fired when the rats moved onto or through a particular spot in the box, just like hippocampal place cells. But they went on to fire at several other spots too. While a rat scurried around mopping up chocolate treats, the researchers watched, perplexed, as the computer mapped the firings, and overlapping blobs appeared on the screen. The Mosers could see that the blobs were creating some sort of pattern, but they couldn't work out what it was.

It took some months before it dawned on them that they needed the rats to run around bigger boxes, so that the pattern would be stretched out and easier to see. At that point, it came into view: a near-perfect hexagon lattice, like a honeycomb. At first they refused to believe it. Such simplicity and regularity was the last thing they had expected — biology is usually a lot messier than this. But one by one, the pair ruled out all other explanations — that the pattern was an artefact from their electronic equipment, for example — and then they began to understand how this part of the brain was working. There were no physical hexagons traced on the floor; the shapes were abstractly created in the rat's brain and imposed on its environment, such that a single neuron fired whenever it crossed one of the points of the hexagon. The discovery was exciting for more than its pleasing pattern. This representation of space in brain-language was one of the long-sought codes by which the brain represents the world around us. "It was a long-drawn-out eureka moment," recalls Edvard. The team published the discovery in *Nature*³ in 2005.

HIDDEN PATTERN

Soon the Mosers were putting the grid cells to the test. They showed that the firing pattern of the cells remained constant even in the dark, and that they are independent of the animals' speed or direction³. Whereas place cells in a rat brain may change their firing rates if their environment is altered even a little — for example by changing the colour of the walls — those of grid cells remain robustly unchanged. The Mosers also found that the different cells in the entorhinal cortex generate grids of many different types, like overlapping honeycombs — big, small and in every orientation and position relative to the box's border. And they ultimately came to see that the brain's grid cells are arranged according to a precise mathematical rule.

The cells that generate smaller grids, with narrower spacing, are at the top of the entorhinal cortex, and those that generate bigger grids are at the bottom. But it is even more exact than that: cells that make grids of the same size and orientation seem to cluster into

modules. The modules are arranged in steps down the length of the entorhinal cortex, and the size of the grid represented by each module expands by a constant factor of 1.4 with every step⁴. At the same time, grid cells that represent different positions relative to the box's border are dotted randomly through the structure. Assuming a similar arrangement exists in humans, the idea is that, together, these cells are unconsciously keeping track of where we are as we wander between rooms or stroll down a street.

ALL IN THE MIND

These discoveries link the Mosers to a rich cast of scientists and philosophers who have pondered the connections between brain, memory and location since at least the time of Ancient Greece. Back then, a philosopher who needed to remember a long speech might memorize the layout of a building or a street, and mentally attach different parts of the speech to its different landmarks. He (they were almost always men) could then fluently deliver the entire rhetoric as he mentally walked around, allowing each landmark to activate the individual sections from memory. The fascination with memory and location continued into the twentieth century, when behavioural scientists first hypothesized that animals carry an abstract map of space inside their heads. The grid cells finally proved that this was true.

The discoveries also astonished and thrilled theoreticians, because the hexagonal pattern is the optimal arrangement for achieving the highest-possible spatial resolution with a minimum number of grid cells. This saves energy, showing how beautifully efficient the brain can sometimes be. "Whoever would have believed that such a beautiful hexagonal representation existed so deep in the brain?" says Andreas Herz, a computational neuroscientist at the University of Munich in Germany. "It was so unexpected that the brain would use the same simple geometric forms that we have been describing in mathematics for millennia." The appealing simplicity gives hope, he says, that the entire brain uses computational principles that scientists may eventually understand.

That understanding could take a long time to reach. It seems unlikely that the neural codes that the brain uses to represent other aspects of the world will be so simple; individual neurons may code for several different properties of the world, making the languages difficult to disentangle. The grid code is also valuable because it exists high up in the brain's hierarchy, with no direct input of sensory information. Unlike the visual cortex, say, whose coding will be influenced by light falling onto the retina, the entorhinal cortex creates the hexagonal pattern entirely internally, by integrating whatever information about the environment is received by other areas of the brain.

With the lab churning out one high-impact paper after another, the Mosers' work has attracted people and funding. Neuroscientist David Rowland was doing his PhD at the University of Oregon in Eugene when he read the 2005 paper on grid cells and was inspired. "I thought it was so cool that I immediately wanted my first post-doc to be in their lab," he says — and that's how it worked out. He has joined the Mosers at the Trondheim Kavli Institute, which is now buzzing with six additional research groups, each working on different aspects of neural circuitry and coding.

Not every couple would find it easy to work together in such apparent harmony. The Mosers ascribe their ability to do so in large part to their patient temperaments and shared interests — in science and beyond. Both love outdoor activities: May-Britt runs every other day across the rugged hills around their coastal home, and Edvard hikes at weekends. They share an obsession with volcanoes — hence their engagement at the top of one — and have climbed many of the globe's most spectacular peaks.

At work, they have evolved some division of labour. Edvard is more involved in computing and theory, and May-Britt manages the lab and staff and is more immersed in the experiments. "We have different strengths and we know that by combining them, the

results become so much better," says Edvard. They aim for only one of them to attend any particular meeting, so that the other is left in the lab. "So we are not really stepping on each other throughout the day, as many people might believe," says Edvard.

The Mosers — and other labs around the world now studying grid cells — still have a lot to learn. Scientists do not yet know how the grid is generated by the neural networks in the entorhinal cortex, or how the overall map created by grid cells, place cells and other navigation cells is integrated to help animals to get from one place to the next. These challenges require more data, and the Mosers have a roster of experiments under way to collect them.

One virtual-reality experiment they are planning will record from

"IT WAS SO UNEXPECTED THAT THE BRAIN WOULD USE THE SAME SIMPLE GEOMETRIC FORMS THAT WE HAVE BEEN DESCRIBING IN MATHEMATICS FOR MILLENNIA."

electrodes in rats running on a stationary ball surrounded by screens showing changing environments. The rats' heads will be held still so that it becomes possible to place electrodes directly inside individual cells for the first time, and to insert small lenses that allow the researchers to simultaneously examine those cells under a microscope. This will reveal precisely which of the many cell types are firing at any one time as the rats move around the virtual space.

The next step will be to map how the grid cells are hard-wired into networks, and to find out when in the rats' lives this wiring happens. Early studies suggest that the grid system is fully established at around three or four weeks after birth^{5,6}, which implies that babies — humans as well as rats — are born with a very primitive sense of where they are in space, and that this sense develops as their brains adapt to the world. The Mosers are also planning to test how the hexagonal pattern would be modified in the brain of a rat that has been reared from birth in a perfect sphere instead of a flat-bottomed cage.

Outside the abstract world of neural coding, grid cells have another major relevance — in understanding memory and its loss. The entorhinal cortex is the first structure in the brain to be affected by Alzheimer's disease, and getting or feeling lost is one of the disease's first symptoms. The Mosers hypothesize that the cells in the entorhinal cortex may have special properties that allow the disease to develop there early — a puzzle that they hope scientists elsewhere can start solving.

Meanwhile, in Trondheim, it is 10 p.m. and Edvard and May-Britt are still discussing the brain as they find their way home. Later, long after they are out of sight, the two scientists continue to make their presence felt. Anyone flying out of Trondheim Airport will find a photograph of the couple in an exhibition of famous Norwegians. The other 13 portraits are all of individual athletes or artists. The Mosers' portrait is the only one featuring two scientific brains. ■ **SEE NEWS P.153**

Alison Abbott is Nature's senior European correspondent.

1. Moser, E., Moser, M.-B. & Andersen, P. *J. Neurosci.* **13**, 3916–3925 (1993).
2. Moser, M.-B. & Moser, E. I. *J. Neurosci.* **18**, 7535–7542 (1998).
3. Hafting, T., Fyhn, M., Molden, S., Moser, M.-B., Moser, E. I. *Nature* **436**, 801–806 (2005).
4. Stensola, H. et al. *Nature* **492**, 72–78 (2012).
5. Langston, R. F. et al. *Science* **328**, 1576–1580 (2010).
6. Wills, T. J., Cacucci, F., Burgess, N. & O'Keefe, J. *Science* **328**, 1573–1576 (2010).

THE BIG REBOOT

As the Large Hadron Collider prepares to come back to life after a two-year hiatus, physicists are gearing up to go beyond the standard model of particle physics.

Mike Lamont grabs the last croissant from a table and eats it as he walks through the control centre at CERN, the European laboratory for particle physics just outside Geneva, Switzerland. It is mid-morning, and the vast blue room is full of physicists staring into computer screens. Lamont, the operations manager of CERN's beams department, explains that they are running tests to ensure that an unexpected computer outage would not affect the network of electronics, vacuum pipes and superconducting magnets that comprise CERN's Large Hadron Collider (LHC), the most powerful particle accelerator in the world.

BY MATTHEW CHALMERS

This is one of numerous checks that are helping Lamont and his colleagues to sleep better at night. They are nearing completion of a major refurbishment that has been under way since March 2013. They have already started cooling down the accelerator's 27-kilometre ring of superconducting magnets in preparation for a restart next year. But when the LHC comes back to life, circulating its twin beams of protons in opposite directions around the ring, Lamont and his colleagues will be pushing to get as close as possible to its design energy of 7 trillion electron volts (TeV) per beam — nearly twice what the LHC has managed so far. Each beam will pack as much energy as a speeding freight train.

Lamont knows all too well what can happen if things go wrong. He was here in September 2008, when the team last attempted to ramp up the US\$5-billion collider to such energies — and ended up triggering an electrical fault that knocked it out of commission for more than a year and cost tens of millions of dollars to repair.

"We've learned a lot about the machine since then," says Lamont. The researchers managed to patch it up and get it working again by the end of 2009, although they ran it at only half its design energy to avoid another shutdown. Still, that was enough for collisions between the beams to produce conclusive evidence for the long-sought Higgs boson — the last unconfirmed prediction of the 40-year-old standard model of particle physics, which describes the behaviour of every particle and force known except gravity.

But for all the acclaim that greeted the announcement of the Higgs particle in July 2012 — and the 2013 Nobel Prize awarded to the theorists who first conjectured its existence — there is much more that the LHC physicists hope to learn from the machine's upcoming run. Is the newly discovered Higgs the only particle of its kind, as the standard model predicts, or is it just the lightest member of a whole family? If there are more Higgs particles, some of them might appear at higher collision energies. Or perhaps the high energies will produce other new, exotic particles that have no place in the standard model.

Theorists have been predicting such particles for decades. Supersymmetry, an extension of the standard model first proposed in the early

1970s, holds that each particle has a heavier counterpart or 'sparticle', and that the two differ in predictable ways.

One or more such sparticles might turn out to be the constituents of dark matter — an invisible haze that is massive enough to control the motion of galaxies, but that is unaccounted for in the standard model. Finding these sparticles, assuming that they are not too heavy to be produced at LHC energies, will thus be a prime goal for the refurbished machine. It might even turn up still more exotic results, such as evidence for spatial dimensions beyond the familiar three. But first, Lamont and his team have to get the LHC running at full capacity.

CLICKS AND HISSES

After a short drive from the control centre, Lamont dons a helmet, steel-capped boots and emergency breathing equipment, then steps into an elevator for a journey 100 metres underground. The elevator doors open onto a service corridor. From there, a short walk leads to the LHC tunnel, where a string of cylindrical, bright-blue magnets curves gently into the distance.

Even after 25 years at CERN, Lamont says that he is still in awe of the power and complexity of the machine. It seems light years away from the cerebral calm of the control room. Down here, the LHC hums, clicks and hisses, and its tunnel smells of metal, dust and warm circuitry. The 15-metre-long, 35-tonne magnets are held off the concrete floor by heavy-duty jacks, and are packed with intricate wiring and plumbing that encases the airtight beam pipes running through their centres. To avert another short circuit, the LHC has been fitted with sensors and thousands of kilometres of cable to detect the slightest sign of a surge in voltage. Crucially, 10,000 of the superconducting connectors that link the magnets have been reinforced or replaced — a task that took more than 250 people more than a year to complete.

Since June, the team has been cooling the magnets down towards their operating temperature of 1.9 kelvin — at which the current-carrying cables that generate the magnetic fields become superconductive. To keep this process manageable, the LHC ring is divided into eight sections that can be refrigerated independently. Once the magnets have been chilled, which takes two months per section, the team will carry out electrical tests to ensure that they can operate at high energy. Lamont already knows that things will not go smoothly. There is a batch of magnets that performed perfectly in tests above ground, but, for some reason, 'quench' or lose their superconductivity when they reach magnetic fields equivalent to beam energies of just 6.5 TeV. This is not a disaster — fixing the magnets is just a matter of cycling each one through several quenches until it settles down and works properly. But it does take time, he says. "And there are hundreds of the buggers!"

Eventually, however, proton beams will once again be threaded



ILLUSTRATION BY ANDY POTTS

through the LHC — a milestone now scheduled for March 2015. And then, after another few weeks of testing, the physicists will start carefully steering the beams into collisions and checking that it is safe for the detectors to start data collection.

There is a faint smell of burning in the tunnel. Lamont explains that a vacuum pipe is being heated to drive off stray molecules. He walks past the magnets to a point where the bare beam line plunges through a massive copper and steel wall. On the other side lies ATLAS, one of the LHC's four main particle detectors (see 'The refurbished ring'). Soon, bunches of high-energy protons will be fired past this point into ATLAS, where they will slam into equally energetic protons circulating in the other direction and send collision debris streaming outwards through the detector. "You think: Jesus, they let us steer a beam through here?" says Lamont. "I still can't believe I get paid to play around with this thing."

SYSTEM UPDATES

Some 8.5 kilometres away from ATLAS, on the opposite side of the LHC ring, Tiziano Camporesi stares up at the 12,500-tonne Compact Muon Solenoid (CMS) — and marvels at the audacity of the physicists who designed it 30 years ago. "They must have been crazy," he says. Many people declared the machine — a vast cylinder with concentric layers of silicon particle sensors, superconducting magnets and massive iron 'yokes' to contain the magnetic field — too intricate to ever work. But it did, says Camporesi, and "far better than we ever expected". CMS and ATLAS were the detectors that identified the Higgs boson in 2012.

"YOU THINK: JESUS, THEY LET US STEER A BEAM THROUGH HERE? I STILL CAN'T BELIEVE I GET TO PLAY AROUND WITH THIS THING."

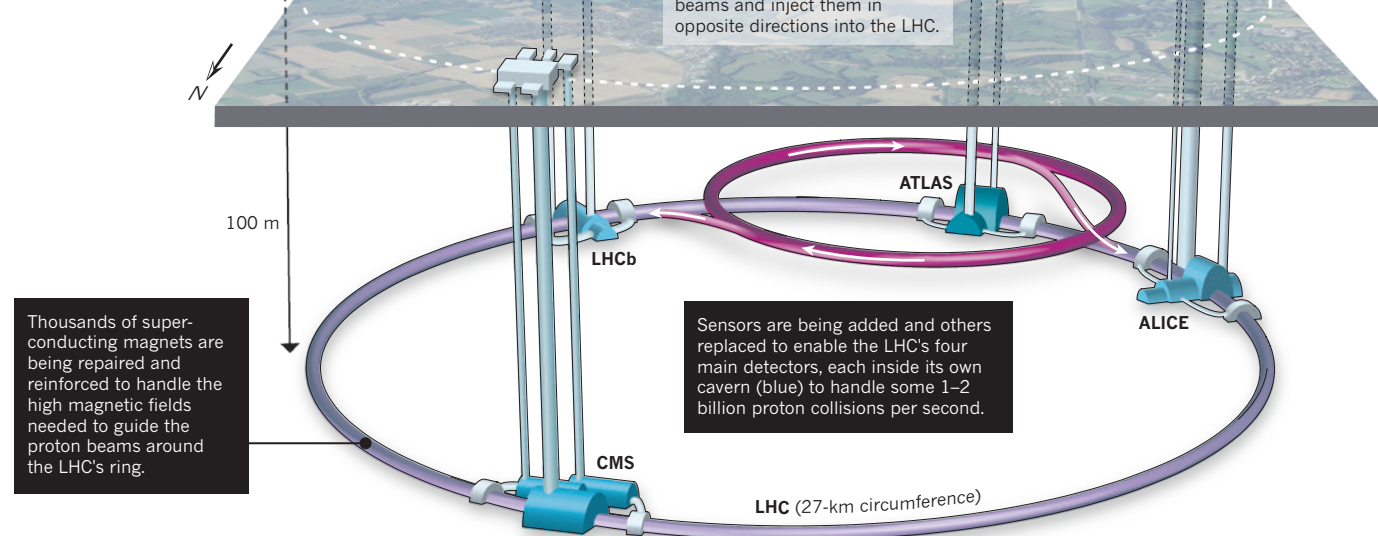
Camporesi, who was elected spokesman of the 3,800-strong CMS collaboration earlier this year, is now coordinating activities for next year's high-energy run. His team, like those on all of the LHC's main experiments (which also include the more specialized ALICE and LHCb detectors located elsewhere around the ring) have been making some much needed repairs and upgrades during the downtime. They have had one piece of good news: in the central region of the detector, where the beams intersect and newly created particles

explode outwards from the collision point, the sensitive silicon trackers have so far survived without radiation damage. But CMS physicists have replaced several faulty photomultiplier tubes that were giving false results, making it seem as though an exotic new particle had been produced when none had.

Camporesi is particularly proud of the four disc-shaped chambers added to each end of the CMS to improve its ability to sense particles called muons. This upgrade, in turn, will beef up the detector's 'trigger' — a combination of electronics and software that will monitor the particles streaming through the detector from the collisions and look for patterns that could signify an event worthy of further study. Physicists have been using such triggers for decades, says Camporesi. But the LHC's next run will not only boost the energy of each beam, it will also increase the number of protons they carry. The result will be some 1–2 billion collisions per second in the CMS. Particles from one collision will still be making their way out of the detector while as many as 50 new collisions occur behind them. From all those events, the trigger

THE REFURBISHED RING

Physicists at the Large Hadron Collider (LHC) near Geneva, Switzerland, have used a 2-year shutdown to make repairs that will boost the accelerator's energy to nearly 7 trillion electron volts per beam.



will have to decide which to store for further analysis; the goal is to bring the final recorded event rate down to several hundred per second. "It's occupying a lot of our time right now," says Camporesi.

BIG DATA CHALLENGE

Once the refurbished LHC is running again, the raw electronic signals from CMS and the other detectors will flow back to the main CERN campus through optical fibres that link directly into the laboratory's computing centre — a stuffy, windowless room in which row upon row of racks hold some 100,000 processors, and cooling fans work noisily to control the heat.

The processors will analyse the incoming data with algorithms that determine the identity, energy and direction of each particle emerging from each collision. The results will then be stored on magnetic tape — an old-fashioned medium that has the advantage of being cheaper and more durable than digital storage.

But just storing the information at CERN would not satisfy researchers' near-insatiable appetite for it. Today's particle physicists spend most of their time writing thousands of lines of computer code designed to search millions of collisions for unusual signals. To get data to these researchers, CERN set up the Worldwide Computing Grid, in which the computer centre streams copies of the data to 13 'tier-1' computer centres worldwide. These centres, in turn, are connected to more than 150 smaller computer clusters called tier-2 nodes, most of which are at universities.

Fortunately for the end users, they do not have to know about any of this. A physicist just has to submit a program to the grid and specify which collision events are to be examined. Grid software will then automatically shunt the job to a centre that has enough processing power and disk space free to run it, then return the results (see *Nature* 469, 282–283; 2011). On this particular day at the CERN computer centre, a live screen shows that 10,500 programs are running at this facility alone — and it represents only 6% of the grid's resources. Were it not for the grid, says Jeremy Coles, a physicist at the University of Cambridge, UK, who is the grid's UK coordinator, his colleagues would probably still be searching for the Higgs boson.

The challenge for the future, says Coles, is dealing with the sky-high event rates to come. During the LHC's first run, despite radical pruning by the detectors' triggers, the data still piled up at the rate of 15 petabytes

(15,000 trillion bytes) a year — more than in all the videos uploaded to YouTube annually. When the LHC starts up again next year, the doubling of the collision rate will bump that up to roughly 30 petabytes per year — an average of about 1 gigabyte per second.

Coles is confident that the grid will be able cope with that increase — not least because of technological advances that have enabled much closer integration between computer centres. "Networking has come on very quickly in the past 10 years," he says — "more than we thought." Just last year, for example, CERN expanded the capability of its data centre, which is at the limits of available space and power, by linking it up with a facility in Budapest through two fibres that transfer data at 100 gigabits per second. From an operations point of view, says Coles, it is just like having the machines in the room next door.

But the data onslaught will not stop there. Planned upgrades to the LHC will send its data output soaring to 110 petabytes a year by the early 2020s, and, eventually, to as much as 400 petabytes a year. "We currently have no way to deal with this," says Coles. Making matters worse, computer chip speeds are stagnating. The best commercial chips now available often contain two, four or eight processors to boost their power. Future chips are likely to have even more. But the code that runs the LHC was written to be run on one processor at a time, says Coles. Finding ways to run data analysis on many processors in parallel would mean rewriting 15 million lines of code written by thousands of people over many years.

Still, when CERN physicists needed a better way to share information in the late 1980s, the result was the World Wide Web. And when they needed a better way to access computer resources back in the 1990s, they invented the world's largest computing grid. The LHC scientists seem confident that they will invent a way around this problem as well.

Lamont seems to feel similarly confident when he talks about the 'next big collider' in particle physics. Although CERN just celebrated its sixtieth anniversary, and the LHC still has another 20 years of proton smashing to go, the laboratory is exploring the feasibility of a collider 80–100 kilometres around that would drill even deeper into the structure of matter (see *Nature* 503, 177; 2013). Lamont says that he would be lucky to see such a machine built during his lifetime, but he points out that the LHC, which came online for the first time in 2008, was first sketched out in 1984. "We've got to start thinking about the next machine now," he says. ■

Matthew Chalmers is a freelance writer in Bristol, UK.

COMMENT

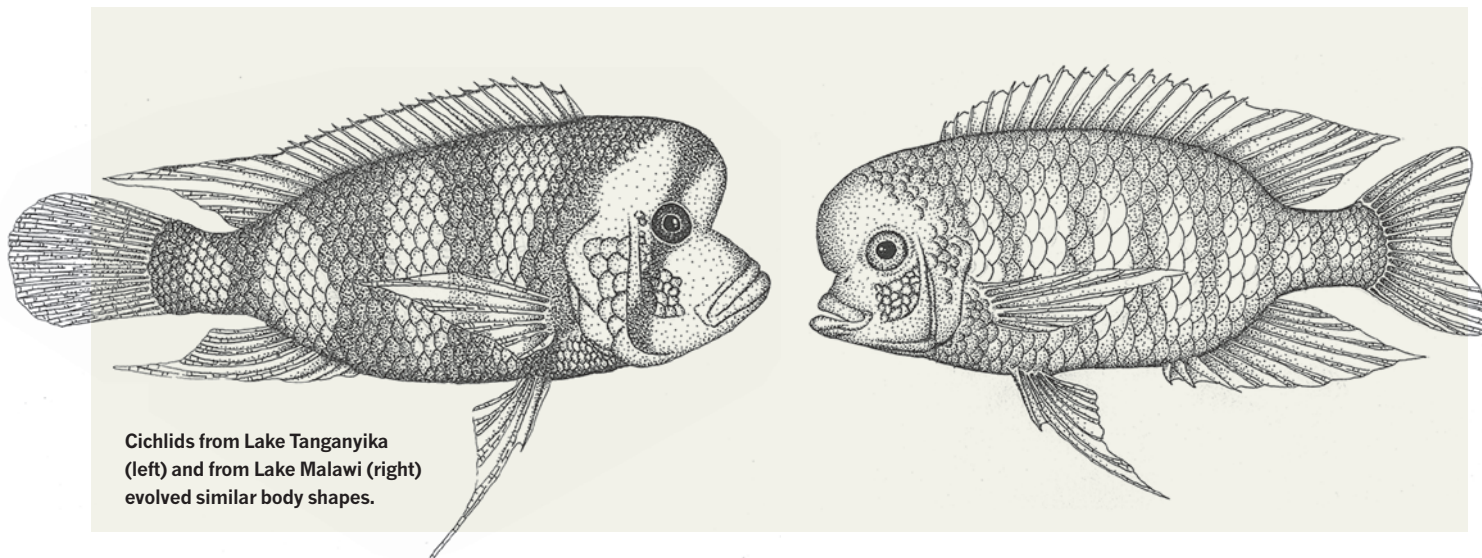
HEALTH Lasting legacy of wartime battle against malaria **p.166**



AGEING Atul Gawande's call to action on end-of-life medical care **p.167**

ENERGY Don't assume that renewable energies are problem-free **p.168**

HISTORY Nobel physicist talks plants with a waiter, then what? **p.168**



Cichlids from Lake Tanganyika (left) and from Lake Malawi (right) evolved similar body shapes.

Does evolutionary theory need a rethink?

Researchers are divided over what processes should be considered fundamental.

POINT

Yes, urgently

Without an extended evolutionary framework, the theory neglects key processes, say Kevin Laland and colleagues.

Charles Darwin conceived of evolution by natural selection without knowing that genes exist. Now mainstream evolutionary theory has come to focus almost exclusively on genetic inheritance and processes that change gene frequencies.

Yet new data pouring out of adjacent fields are starting to undermine this narrow stance. An alternative vision of evolution is beginning to crystallize, in which the processes by which organisms grow and develop are recognized as causes of evolution.

Some of us first met to discuss these advances six years ago. In the time since, as members of an interdisciplinary team, we have worked intensively to develop a broader framework, termed the extended evolutionary synthesis¹ (EES), and to flesh out its structure, assumptions and predictions. In essence, this synthesis maintains that important drivers of evolution, ones that cannot be reduced to genes, must be woven into the very fabric of evolutionary theory.

We believe that the EES will shed new light on how **PAGE 162 ►**

COUNTERPOINT

No, all is well

Theory accommodates evidence through relentless synthesis, say Gregory A. Wray, Hopi E. Hoekstra and colleagues.

In October 1881, just six months before he died, Charles Darwin published his final book. *The Formation of Vegetable Mould, Through the Actions of Worms*¹¹ sold briskly: Darwin's earlier publications had secured his reputation. He devoted an entire book to these humble creatures in part because they exemplify an interesting feedback process: earthworms are adapted to thrive in an environment that they modify through their own activities.

Darwin learned about earthworms from conversations with gardeners and his own simple experiments. He had a genius for distilling penetrating insights about evolutionary processes — often after amassing years of observational and experimental data — and he drew on such disparate topics as agriculture, geology, embryology and behaviour. Evolutionary thinking ever since has followed Darwin's lead in its emphasis on evidence and in synthesizing information from other fields.

A profound shift in evolutionary thinking began **PAGE 163 ►**

ILLUSTRATION BY R. CRAIG ALBERTSON

POINT: YES, URGENTLY ▶ evolution works. We hold that organisms are constructed in development, not simply ‘programmed’ to develop by genes. Living things do not evolve to fit into pre-existing environments, but co-construct and coevolve with their environments, in the process changing the structure of ecosystems.

The number of biologists calling for change in how evolution is conceptualized is growing rapidly. Strong support comes from allied disciplines, particularly developmental biology, but also genomics, epigenetics, ecology and social science^{1,2}. We contend that evolutionary biology needs revision if it is to benefit fully from these other disciplines. The data supporting our position gets stronger every day.

Yet the mere mention of the EES often evokes an emotional, even hostile, reaction among evolutionary biologists. Too often, vital discussions descend into acrimony, with accusations of muddle or misrepresentation. Perhaps haunted by the spectre of intelligent design, evolutionary biologists wish to show a united front to those hostile to science. Some might fear that they will receive less funding and recognition if outsiders — such as physiologists or developmental biologists — flood into their field.

However, another factor is more important: many conventional evolutionary biologists study the processes that we claim are neglected, but they comprehend them very differently (see ‘No, all is well’). This is no storm in an academic tearoom, it is a struggle for the very soul of the discipline.

Here we articulate the logic of the EES in the hope of taking some heat out of this debate and encouraging open discussion of the fundamental causes of evolutionary change (see Supplementary Information; go.nature.com/boffk7).

CORE VALUES

The core of current evolutionary theory was forged in the 1930s and 1940s. It combined natural selection, genetics and other fields into a consensus about how evolution occurs. This ‘modern synthesis’ allowed the evolutionary process to be described mathematically as frequencies of genetic variants in a population change over time — as, for instance, in the spread of genetic resistance to the myxoma virus in rabbits.

In the decades since, evolutionary biology has incorporated developments consistent with the tenets of the modern synthesis. One such is ‘neutral theory’, which emphasizes random events in evolution. However, standard evolutionary theory (SET) largely retains the same assumptions as the original modern synthesis, which continues to channel how people think about evolution.

The story that SET tells is simple: new variation arises through random genetic mutation; inheritance occurs through DNA; and natural selection is the sole cause of adaptation, the process by which organisms become well-suited to their environments. In this view, the complexity of biological development — the changes that occur as an organism grows and ages — are of secondary, even minor, importance.

In our view, this ‘gene-centric’ focus fails to capture the full gamut of processes that direct evolution. Missing pieces include how physical development influences the generation of variation (developmental bias); how the environment directly shapes organisms’ traits (plasticity); how organisms modify environments (niche construction); and how organisms transmit more than genes across generations (extra-genetic inheritance). For SET, these phenomena are just outcomes of evolution. For the EES, they are also causes.

Valuable insight into the causes of adaptation and the appearance of new traits comes from the field of evolutionary developmental biology (‘evo-devo’). Some of its experimental findings are proving tricky to assimilate into SET. Particularly thorny is the observation that much variation is not random because developmental processes generate certain forms more readily than others³. For example, among

one group of centipedes, each of the more than 1,000 species has an odd number of leg-bearing segments, because of the mechanisms of segment development³.

In our view, this concept — developmental bias — helps to explain how organisms adapt to their environments and diversify into many different species. For example, cichlid fishes in Lake Malawi are more closely related to other cichlids in Lake Malawi than to those in Lake Tanganyika, but species in both lakes have strikingly similar body shapes⁴. In each case, some fish have large fleshy lips, others protruding foreheads, and still others short, robust lower jaws.

SET explains such parallels as convergent evolution: similar environmental conditions select for random genetic variation with equivalent results. This account requires extraordinary coincidence to explain the multiple parallel forms that evolved independently in each lake. A

more succinct hypothesis is that developmental bias and natural selection work together^{4,5}. Rather than selection being free to traverse across any physical possibility, it is guided along specific routes opened up by the processes of development^{5,6}.

Another kind of developmental bias occurs when individuals respond to their environment by changing their form — a phenomenon called plasticity. For instance, leaf shape changes with soil water and chemistry. SET views this plasticity as merely fine-tuning, or even noise. The EES sees it as a plausible first step in adaptive evolution. The key finding here is that plasticity not only allows organisms to cope in new environmental conditions but to generate traits

that are well-suited to them. If selection preserves genetic variants that respond effectively when conditions change, then adaptation largely occurs by accumulation of genetic variations that stabilize a trait after its first appearance^{5,6}. In other words, often it is the trait that comes first; genes that cement it follow, sometimes several generations later⁷.

Studies of fish, birds, amphibians and insects suggest that adaptations that were, initially, environmentally induced may promote colonization of new environments and facilitate speciation^{5,6}. Some of the best-studied examples of this are in fishes, such as sticklebacks and Arctic char. Differences in the diets and conditions of fish living at the bottom and in open water have induced distinct body forms, which seem to be evolving reproductive isolation, a stage in forming new species. The number of species in a lineage does not depend solely on how random genetic variation is winnowed through different environmental sieves. It also hangs on developmental properties that contribute to the lineage’s ‘evolvability’.

In essence, SET treats the environment as a ‘background condition’, which may trigger or modify selection, but is not itself part of the evolutionary process. It does not differentiate between how termites become adapted to mounds that they construct and, say, how organisms adapt to volcanic eruptions. We view these cases as fundamentally different⁷.

Volcanic eruptions are idiosyncratic events, independent of organisms’ actions. By contrast, termites construct and regulate their homes in a repeatable, directional manner that is shaped by past selection and that instigates future selection. Similarly, mammals, birds and insects defend, maintain and improve their nests — adaptive responses to nest building that have evolved again and again⁷. This ‘niche construction’, like developmental bias, means that organisms co-direct their own evolution by systematically changing environments and thereby biasing selection⁷.

INHERITANCE BEYOND GENES

SET has long regarded inheritance mechanisms outside genes as special cases; human culture being the prime example. The EES explicitly recognizes that parent–offspring similarities result in part from parents reconstructing their own developmental environments for their offspring. ‘Extra-genetic inheritance’ includes **PAGE 164** ▶



Plasticity: commodore butterflies emerge with different colours in dry (left) and wet seasons.

ORANGE: PETER CHADWICK/SPL BLUE: LAWRENCE LAWRY/SPL

COUNTERPOINT: NO, ALL IS WELL ▶ during the 1920s, when a handful of statisticians and geneticists began quietly laying the foundations for a dramatic transformation. Their work between 1936 and 1947 culminated in the ‘modern synthesis’, which united Darwin’s concept of natural selection with the nascent field of genetics and, to a lesser extent, palaeontology and systematics. Most importantly, it laid the theoretical foundations for a quantitative and rigorous understanding of adaptation and speciation, two of the most fundamental evolutionary processes.

In the decades since, generations of evolutionary biologists have modified, corrected and extended the framework of the modern synthesis in countless ways. Like Darwin, they have drawn heavily from other fields. When molecular biologists identified DNA as the material basis for heredity and trait variation, for instance, their discoveries catalysed fundamental extensions to evolutionary theory. For example, the realization that many genetic changes have no fitness consequences led to major theoretical advances in population genetics. The discovery of ‘selfish’ DNA prompted discussions about selection at the level of genes rather than traits. Kin selection theory, which describes how traits affecting relatives are selected, represents another extension¹².

Nonetheless there are evolutionary biologists (see ‘Yes, urgently’) who argue that theory has since ossified around genetic concepts. More specifically, they contend that four phenomena are important evolutionary processes: phenotypic plasticity, niche construction, inclusive inheritance and developmental bias. We could not agree more. We study them ourselves.

But we do not think that these processes deserve such special attention as to merit a new name such as ‘extended evolutionary synthesis’. Below we outline three reasons why we believe that these topics already receive their due in current evolutionary theory.

NEW WORDS, OLD CONCEPTS

The evolutionary phenomena championed by Laland and colleagues are already well integrated into evolutionary biology, where they have long provided useful insights. Indeed, all of these concepts date back to Darwin himself, as exemplified by his analysis of the feedback that occurred as earthworms became adapted to their life in soil.

Today we call such a process niche construction, but the new name does not alter the fact that evolutionary biologists have been studying feedback between organisms and the environment for well over a century¹³. Stunning adaptations such as termite mounds, beaver dams, and bowerbird displays have long been a staple of evolutionary studies. No less spectacular are cases that can only be appreciated at the microscopic or molecular scale, such as viruses that hijack host cells to reproduce and ‘quorum sensing’, a sort of group think by bacteria.

Another process, phenotypic plasticity, has drawn considerable attention from evolutionary biologists. Countless cases in which the environment influences trait variation have been documented — from the jaws of cichlid fishes that change shape when food sources alter,

to leaf-mimicking insects that are brown if born in the dry season and green in the wet. Technological advances in the past decade have revealed an incredible degree of plasticity in gene expression in response to diverse environmental conditions, opening the door to understanding its material basis. Much discussed, too, was a book⁵ by behavioural scientist Mary Jane West-Eberhard that explored how plasticity might precede genetic changes during adaptation.

So, none of the phenomena championed by Laland and colleagues are neglected in evolutionary biology. Like all ideas, however, they need to prove their value in the marketplace of rigorous theory, empirical results and critical discussion. The prominence that these four phenomena command in the discourse of contemporary evolutionary theory reflects their proven explanatory power, not a lack of attention.

MODERN EXPANSION

Furthermore, the phenomena that interest Laland and colleagues are just four among many that offer promise for future advances in evolutionary biology. Most evolutionary biologists have a list of topics that they would like to see given more attention. Some would argue that epistasis — complex interactions among genetic variants — has long been under-appreciated. Others would advocate for cryptic genetic variation (mutations that affect only traits under specific genetic or environmental conditions). Still others would stress the importance of extinction, or adaptation to climate change, or the evolution of behaviour. The list goes on.

We could stop and argue about whether ‘enough’ attention is being paid to any of these. Or we could roll up our sleeves, get to work, and find out by laying the theoretical foundations and building a solid casebook of empirical studies. Advocacy can take an idea only so far.

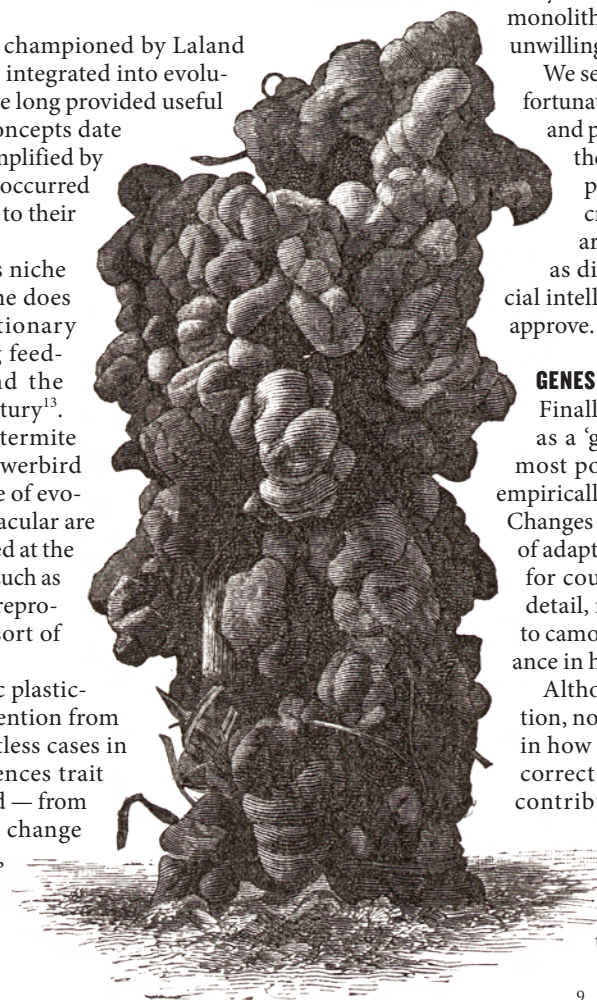
What Laland and colleagues term the standard evolutionary theory is a caricature that views the field as static and monolithic. They see today’s evolutionary biologists as unwilling to consider ideas that challenge convention.

We see a very different world. We consider ourselves fortunate to live and work in the most exciting, inclusive and progressive period of evolutionary research since the modern synthesis. Far from being stuck in the past, current evolutionary theory is vibrantly creative and rapidly growing in scope. Evolutionary biologists today draw inspiration from fields as diverse as genomics, medicine, ecology, artificial intelligence and robotics. We think Darwin would approve.

GENES ARE CENTRAL

Finally, diluting what Laland and colleagues deride as a ‘gene-centric’ view would de-emphasize the most powerfully predictive, broadly applicable and empirically validated component of evolutionary theory. Changes in the hereditary material are an essential part of adaptation and speciation. The precise genetic basis for countless adaptations has been documented in detail, ranging from antibiotic resistance in bacteria to camouflage coloration in deer mice, to lactose tolerance in humans.

Although genetic changes are required for adaptation, non-genetic processes can sometimes play a part in how organisms evolve. Laland and colleagues are correct that phenotypic plasticity, for instance, may contribute to the adaptedness of an individual. A seedling might bend towards brighter light, growing into a tree with a different shape from its siblings. Many studies have shown that this kind of plasticity is beneficial, and that it can readily evolve if there



A worm cast pictured in Charles Darwin’s final book.

POINT: YES, URGENTLY ▶ the transmission of epigenetic marks (chemical changes that alter DNA expression but not the underlying sequence) that influence fertility, longevity and disease resistance across taxa⁸. In addition, extra-genetic inheritance includes socially transmitted behaviour in animals, such as nut cracking in chimpanzees or the migratory patterns of reef fishes^{8,9}. It also encompasses those structures and altered conditions that organisms leave to their descendants through their niche construction — from beavers' dams to worm-processed soils^{7,10}. Research over the past decade has established such inheritance to be so widespread that it should be part of general theory.

Mathematical models of evolutionary dynamics that incorporate extra-genetic inheritance make different predictions from those that do not⁷⁻⁹. Inclusive models help to explain a wide range of puzzling phenomena, such as the rapid colonization of North America by the house finch, the adaptive potential of invasive plants with low genetic diversity, and how reproductive isolation is established.

Such legacies can even generate macro-evolutionary patterns. For instance, evidence suggests that sponges oxygenated the ocean and by doing so created opportunities for other organisms to live on the seabed¹⁰. Accumulating fossil data indicate that inherited modifications of the environment by species has repeatedly facilitated, sometimes after millions of years, the evolution of new species and ecosystems¹⁰.

BETTER TOGETHER

The above insights derive from different fields, but fit together with surprising coherence. They show that variation is not random, that there is more to inheritance than genes, and that there are multiple routes to the fit between organisms and environments. Importantly, they demonstrate that development is a direct cause of why and how adaptation and speciation occur, and of the rates and patterns of evolutionary change.

SET consistently frames these phenomena in a way that undermines their significance. For instance, developmental bias is generally taken to impose 'constraints' on what selection can achieve — a hindrance that explains only the absence of adaptation. By contrast, the EES recognizes developmental processes as a creative element, demarcating which forms and features evolve, and hence accounting for why organisms possess the characters that they do.

Researchers in fields from physiology and ecology to anthropology are running up against the limiting assumptions of the standard evolutionary framework without realizing that others are doing the same. We believe that a plurality of perspectives in science encourages development of alternative hypotheses, and stimulates empirical work. No longer a protest movement, the EES is now a credible framework inspiring useful work by bringing diverse researchers under one theoretical roof to effect conceptual change in evolutionary biology. ■

Kevin Laland is professor of behavioural and evolutionary biology at the University of St Andrews, UK. **Tobias Uller, Marc Feldman, Kim Sterelny, Gerd B. Müller, Armin Moczek, Eva Jablonka, John Odling-Smee.**
e-mail: knl1@st-andrews.ac.uk

- Pigliucci, M. & Müller, G. B. *Evolution: The Extended Synthesis* (MIT Press, 2010).
- Noble, D. et al. *J. Physiol.* **592**, 2237–2244 (2014).
- Arthur, W. *Biased Embryos and Evolution* (Cambridge Univ. Press, 2003).
- Brakefield, P. M. *Trends Ecol. Evol.* **21**, 362–368 (2006).
- West-Eberhard, M. J. *Developmental Plasticity and Evolution* (Oxford Univ. Press, 2003).
- Pfennig D. W. et al. *Trends Ecol. Evol.* **25**, 459–467 (2010).
- Odling-Smee, F. J., Laland, K. N. & Feldman, M. W. *Niche Construction: The Neglected Process in Evolution* (Princeton Univ. Press, 2003).
- Jablonka, E. & Lamb, M. *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life* (MIT Press, 2014).
- Hoppitt, W. & Laland, K. N. *Social Learning: An Introduction to Mechanisms, Methods, and Models* (Princeton Univ. Press, 2013).
- Erwin, D. H. & Valentine, J. W. *The Cambrian Explosion: The Construction of Animal Biodiversity* (Roberts, 2013).

COUNTERPOINT: NO, ALL IS WELL ▶ is genetic variation in the response¹⁴. This role for plasticity in evolutionary change is so well documented that there is no need for special advocacy.

Much less clear is whether plasticity can 'lead' genetic variation during adaptation. More than half a century ago, developmental biologist Conrad Waddington described a process that he called genetic assimilation¹⁵. Here, new mutations can sometimes convert a plastic trait into one that develops even without the specific environmental condition that originally induced it. Few cases have been documented outside of the laboratory, however. Whether this is owing to a lack of serious attention or whether it reflects a genuine rarity in nature can be answered only by further study.

Lack of evidence also makes it difficult to evaluate the role that developmental bias may have in the evolution (or lack of evolution) of adaptive traits. Developmental processes, based on features of the genome that may be specific to a particular group of organisms, certainly can influence the range of traits that natural selection can act on. However, what matters ultimately is not the extent of trait variation, nor even its precise mechanistic causes. What matters is the heritable differences in traits, especially those that bestow some selective advantage. Likewise, there is little evidence for the role of inherited epigenetic modification (part of what was termed 'inclusive inheritance') in adaptation: we know of no case in which a new trait has been shown to have a strictly epigenetic basis divorced from gene sequence. On both topics, further research will be valuable.

All four phenomena that Laland and colleagues promote are 'add-ons' to the basic processes that produce evolutionary change: natural selection, drift, mutation, recombination and gene flow. None of these additions is essential for evolution, but they can alter the process under certain circumstances. For this reason they are eminently worthy of study.

We invite Laland and colleagues to join us in a more expansive extension, rather than imagining divisions that do not exist.

We appreciate their ideas as an important part of what evolutionary theory might become in the future. We, too, want an extended evolutionary synthesis, but for us, these words are lowercase because this is how our field has always advanced¹⁶.

The best way to elevate the prominence of genuinely interesting phenomena such as phenotypic plasticity, inclusive inheritance, niche construction and developmental bias (and many, many others) is to strengthen the evidence for their importance.

Before claiming that earthworms "have played a more important part in the history of the world than most persons would at first suppose"¹¹, Darwin collected more than 40 years of data. Even then, he published only for fear that he would soon be "joining them"¹⁷. ■

Gregory A. Wray is professor of biology at Duke University in Durham, North Carolina, USA. **Hopi E. Hoekstra** is professor of biology at Harvard University in Cambridge, Massachusetts, USA. **Douglas J. Futuyma, Richard E. Lenski, Trudy F. C. Mackay, Dolph Schluter, Joan E. Strassmann.**
e-mails: gwray@duke.edu; hoekstra@oeb.harvard.edu

- Darwin, C. *The Formation of Vegetable Mould, Through the Actions of Worms* (John Murray, 1881).
- Alcock, J. *The Triumph of Sociobiology* (Oxford Univ. Press, 2001).
- Bailey, N. W. *Trends Ecol. Evol.* **27**, 561–569 (2012).
- Wada, H. & Sewall, K. B. *Integ. Comp. Biol.* <http://dx.doi.org/10.1093/icb/icu097> (2014).
- Waddington, C. H. *Nature* **150**, 563–565 (1942).
- Callebaut, W. in *Evolution: The Extended Synthesis* (Pigliucci, M. & Müller, G. B. eds) 443–482 (MIT Press, 2010).
- Browne, J. *Charles Darwin: The Power of Place* Vol. II 479 (Jonathan Cape, 2003).

Full author affiliations accompany these articles online at go.nature.com/boffk7.



During the Second World War, malaria control and prevention was a big problem for the US military.

PUBLIC HEALTH

The malaria wars

Robert Seder applauds the chronicle of a secret US wartime project to vanquish the disease.

Malaria has a long history and a deadly present: every year it kills as many as 1 million people, mostly children in sub-Saharan Africa. Over the past decade, mortality from the disease — caused by the *Plasmodium* parasite transmitted by mosquitoes — has been cut through the use of bed nets and other public-health measures championed by the World Health Organization's Roll Back Malaria campaign, among others.

The effectiveness of such measures may have plateaued, however, and the disease is becoming resistant to drugs such as artemisinin-based therapies. In 2015, the first vaccine (GlaxoSmithKline's RTS,S) will be considered for approval in Africa. It confers 30–50% protection that wanes over time. Given all that, insights into historical grappling with the disease are highly useful.

Such insights abound in *The Malaria Project*, science journalist Karen Masterson's chronicle of a secret Second World War scheme run by the White House and US War Department. In the 1940s, the project drew in leading scientists and clinicians to find a cure for malaria — at the time decimating US troops — and isolated the safe, effective drug chloroquine. As Masterson shows, today's fight against the disease, heavily weighted towards drug and vaccine

development, bears the project's stamp.

Masterson discovered the project in 2004 in a letter in the US National Archives. Writing in 1943, physician George Carden described a plan to test malaria drugs on patients at a psychiatric hospital in Boston, Massachusetts. Carden was part of the Malaria Project, working with chemists, entomologists, immunologists and clinicians to assess some 14,000 malaria-drug candidates.

Masterson relates how scientific understanding of malaria advanced. The project's trailblazing tropical-disease specialist Lowell Coggeshall, for instance, had worked on mosquito control in canal and dam construction since the 1920s. His observations, such as that a cure does not prevent reinfection, remain pertinent. But it was Swedish renal specialist Alf Alving who pinned chloroquine down.

Malaria could not be grown in culture, so the drugs were tested in human volunteers, including conscientious objectors and



The Malaria Project: The U.S. Government's Secret Mission to Find a Miracle Cure
KAREN M. MASTERSON
New American Library: 2014.

medical students. The approach was modelled on 1917 research by Julius Wagner-Jauregg. The Austrian psychiatrist discovered that some people with advanced syphilis (who would otherwise have died) could be cured if they received blood taken from people infected with malaria.

In the Malaria Project's crucial studies, at Stateville Prison in Illinois, Alving recruited prisoners to act as both assistants and test subjects for compounds that were safe and protective in animal tests. He lobbied for inmates who tested the drugs to get higher compensation and certificates of merit for use at parole hearings, and helped them to get jobs after release. One of Alving's assistants was the infamous murderer Nathan Leopold.

Alving's breakthrough came from studying SN-7618, a compound that could be given once a week with minimal toxicity. Ironically, the original version, sontochnin, was made in Nazi-occupied Paris and conveyed to the US military by Vichy medics. Alving found that changing the dosage and intervals at which it was given would cure malaria, and chloroquine became the standard treatment until resistance developed in certain strains. Human volunteers still test malaria vaccines because such safe, effective drugs can cure them of the controlled infection.

A revelation of *The Malaria Project* is the toll malaria took on troops: some half a million US soldiers were infected during the Second World War, for instance. Douglas MacArthur, leading Allied forces against Japan, realized that malaria was weakening troops and launched control operations. Leading troops in Italy, by contrast, George Patton showed little compassion for infected soldiers and did not act — even as Adolf Hitler flooded land to encourage mosquito breeding.

The Malaria Project also covers the key role of the US Rockefeller Foundation and Rockefeller Institute for Medical Research (now Rockefeller University). Rather than search for magic bullets, they focused on malaria eradication through public health and cost-effectiveness. Today, the Bill & Melinda Gates Foundation (from which I receive some funding) plans to use public-health measures as well as drugs and possibly vaccines in its malaria eradication strategy.

Mosquito control, improved education and economic development remain central. But more than 60 years after Alving's breakthrough, the most crucial advances will come from safe, effective and durable vaccines to prevent malaria, and drug treatments to eradicate it. ■

Robert Seder is chief of cellular immunology at the Vaccine Research Center of the US National Institutes of Health in Bethesda, Maryland.
e-mail: rseder@mail.nih.gov

Correspondence

Beyond sharing Earth observations

To improve the accuracy of products derived from shared satellite observations of Earth (see M. A. Wulder and N. C. Coops *Nature* **513**, 30–31; 2014), governments and research institutes also need to share calibration and validation data. Such data are measured on the ground or interpreted from high-resolution satellite imagery.

Satellite images are now available at very high resolution. Crowd-sourced calibration and validation data would vastly improve classification algorithms as well as the accuracy of land-cover products.

Enhanced accuracy would enable remote sensing to be better used for monitoring biodiversity loss and ecosystem dynamics, for example, and for other applications that depend on baseline and changing land cover.

Efforts such as the Global Observation of Forest Cover and Land Dynamics and Geo-Wiki.org are working in this direction. More initiatives are needed to unlock the incredible amount of data that remain confined within institutes and agencies.

Linda See, Steffen Fritz, Ian McCallum *International Institute for Applied Systems Analysis, Laxenburg, Austria.*
see@iiasa.ac.at

Renewables: costly long-distance power

In my view, the long-distance transmission of intermittent renewable electricity is not a cost-effective way to reduce emissions (J. A. Mathews and H. Tan *Nature* **513**, 166–168; 2014). Power transmission is highly inefficient, with 3.5% or more being lost from an ultra-high-voltage line over 1,000 kilometres (see go.nature.com/dli4we).

China can produce renewables cheaply because of economies of scale, market expansion,

low labour costs and minimal environmental regulation. But this may not be sustainable.

Already, the availability of resources and competitive utilization are challenging.

Electricity demands vary in different regions, from low in the wind-rich north and west of the country, to high in coastal areas. The cost of transmitting electricity over such vast distances is unlikely to be offset by increasing wind-turbine operating hours.

The remote use of renewable electricity might be much more expensive per unit than electricity produced by conventional generators, even considering the social cost of carbon. In most cases, the costs will also be higher compared with locally generated renewable energy, even when those resources are limited.

Shuwei Zhang *Draworld Environment Research Center, Beijing, China.*
shuwei.zh@gmail.com

Renewables: can harm environment

The large-scale generation of renewable energy (J. A. Mathews and H. Tan *Nature* **513**, 166–168; 2014) can damage the environment.

In China, for instance, giant hydropower projects include 25 hydropower stations within 100 kilometres of one another on the Jinsha River, and 90 stations on 66 tributaries of the Salween River. These stations pose a large risk to local and regional hydrology, geology and ecology (see, for example, *Nature* **513**, 154–155; 2014).

Conversion of wind kinetic energy into electricity generates noise, affecting local communities and migratory birds, and even local weather and climate (L. Zhou *et al.* *Nature Clim. Change* **2**, 539–543; 2012). Furthermore, the manufacture of solar photovoltaic products can cause serious environmental pollution (see H. Yang *et al.*

Nature **509**, 563; 2014).

Development of renewable-energy projects should always take into account the long-term effects on the local natural, social and technological conditions.

Xin Miao *Harbin Institute of Technology, Harbin, China.*
xin.miao@aliyun.com

Renewables: the sky is the limit

The International Energy Agency (IEA) claims in a report that up to 45% of power generated annually could come from renewable energy sources in some advanced economies without significantly increasing long-term power-system costs (*The Power of Transformation*; IEA, 2014). This raises the possibility that certain regions could be using 100% renewables during our lifetime (see also J. A. Mathews and H. Tan *Nature* **513**, 166–168; 2014).

This possibility is backed, for instance, by analyses in India by the conservation group WWF and the Energy and Resources Institute (see go.nature.com/7shcug), and in Australia by the Australian Energy Market Operator (see go.nature.com/wtt1ps).

As noted in the IEA report, progress in integrating renewables depends on cost constraints, local weather and daylight patterns, and the flexibility of existing power systems.

Don Gunasekera *Victoria University, Melbourne, Australia.*
don.gunasekera@vu.edu.au

Photosynthesis and the Nobel physicist

It is not widely known that the Nobel-prizewinning atomic physicist James Franck (see ‘50 Years Ago’ *Nature* **512**, 381; 2014) spent 30 years investigating photosynthesis (see J. L. Rosenberg *Photosynth. Res.* **80**, 71–76; 2004).

Dining at the University of Chicago’s Quadrangle Club, Franck met the distinguished

scientist and *Biochemistry* textbook author Lubert Stryer, then a student and part-time waiter. Stryer told me that Franck chatted to him about his research into how plants convert visible light into chemical energy, adding that it might be useful to Stryer one day.

Later, Stryer took advantage of Franck’s unfinished work to investigate other photochemical reactions, notably going on to elucidate the function of the G protein that transduces the light signal in visual excitation (see L. Stryer *J. Biol. Chem.* **287**, 15164–15173; 2012).

Min-Liang Wong *National Chung-Hsing University, Taiwan.*
mlwong@dragon.nchu.edu.tw

Maths medal boosts Brazilian morale

This year has been momentous for Brazil. We lost the World Cup on our own turf, yet one of our mathematicians received the highest scientific honour ever to be awarded to a Brazilian. Scientists now stand a chance of competing with footballers as role models for Brazil’s youngsters.

Artur Avila, who also has French nationality, won the Fields Medal, the most prestigious prize in maths (see *Nature* <http://doi.org/vn4>; 2014). This triumph has sent a wave of motivation through students and boosted the morale of educators and scientists. Science, as well as football, can offer a route out of poverty.

Tiago Campos Pereira *University of São Paulo, Brazil.*
tiagocampospereira@ffclrp.usp.br

CORRECTION

The Outlook article ‘Early warning system’ (*Nature* **513**, S4–S6; 2014) wrongly said that a low-dose CT scan gives the same radiation exposure as one transatlantic flight. In fact, radiation on such a flight is ~50 microsieverts, 30 times less than from a low-dose scan.

POLRMT does not transcribe nuclear genes

ARISING FROM J. E. Kravchenko, I. B. Rogozin, E. V. Koonin & P. M. Chumakov *Nature* **436**, 735–739 (2005); doi:10.1038/nature03848

Mitochondria are involved in a variety of metabolic processes and one of their main functions is to perform oxidative phosphorylation^{1,2}, which requires a crosstalk between the mitochondrial and nuclear genomes to accomplish coordinated gene expression^{3,4}. Splice variants of the mitochondrial RNA polymerase gene (*Polrmt*) have been reported to encode a nuclear RNA polymerase isoform (spRNAP-IV), which is thought to facilitate this coordination by transcribing a specific subset of nuclear genes^{5–7}. Here we report that analysis of *Polrmt* gene expression, subcellular fractionation and fluorescence microscopy do not support the existence of a nuclear POLRMT isoform in mouse and human cells, and that conditional knockout of *Polrmt* does not affect expression of the nuclear genes previously reported to be transcribed by spRNAP-IV. We thus conclude that POLRMT has an exclusive mitochondrial role and that it is absolutely required for expression of mitochondrial DNA (mtDNA) in mammalian mitochondria.

Alternative splicing of *Polrmt* transcripts has been reported to produce a mitochondrial and a nuclear RNA polymerase isoform denoted POLRMT and spRNAP-IV, respectively^{5,7}. We generated conditional *Polrmt* knockout mice by targeting of exon 3 to disrupt the expression of both protein isoforms. Western blot analyses confirmed that the POLRMT protein was absent (Fig. 1a), leading to a profound reduction of *de novo* transcription of mtDNA (Fig. 1b) in hearts of conditional knockout mice. Next, we used real-time quantitative reverse transcription PCR (qRT-PCR) to assess the expression of the nuclear genes previously reported to be transcribed by spRNAP-IV^{5,6} (that is, muscle actin genes (*Actc1*, *Acta1*, *Actg1*), the zinc-finger BTB domain-containing protein 1 gene (*Zbtb1*) and the prenylcysteine oxidase gene (*Pcyox1*)) and found no significant decrease in levels of these transcripts in *Polrmt* knockout hearts (Fig. 1c) or skeletal muscle (Appendix Fig. 1). The expression of sarcoplasmic/endoplasmic reticulum calcium ATPase2a (*Serca2a*) was reduced in heart (Fig. 1c), as typically seen in cardiomyopathy⁸. On the basis of these findings, we performed reverse transcription PCR (RT-PCR) analyses to verify the existence of the previously described alternatively spliced *Polrmt* transcripts, containing intron 1 or intron 2 sequences, in the mouse^{5,7}. Unexpectedly, these previously reported mRNA isoforms were not detected in heart, muscle, liver, kidney or brain, despite that our assay robustly amplifies corresponding sequences from genomic DNA (Fig. 1d and Appendix Fig. 2). Next, we performed western blot analyses with an antibody directed against a peptide in the carboxy terminus of mouse POLRMT, but could not find the previously reported⁵ nuclear protein isoform of ~110 kDa in mouse heart, muscle, kidney, spleen, liver or brain, whereas a ~140-kDa protein corresponding to mitochondrially localized POLRMT was present in all tissues (Fig. 2a). We also studied the subcellular localization of POLRMT in mouse heart, and in HeLa, 143B and 143B rho⁰ human cell lines and found no nuclear spRNAP-IV isoform on western blot analysis (Fig. 2b and Appendix Fig. 3) and no nuclear signal on confocal microscopy ($n = 45$ cells; Fig. 2c and Appendix Fig. 4) with a POLRMT antibody recognizing both protein isoforms (Appendix Fig. 5). In addition, we expressed spRNAP-IV fused to green fluorescent protein (EGFP) or a Flag tag at its C terminus in HeLa cells and found a strong, predominantly extranuclear localization ($n = 18$; Fig. 2d and Appendix Figs 6 and 7). Moreover, we could not verify the existence of the reported alternatively spliced POLRMT transcript, containing intron 1 sequences^{5,7}, when performing RT-PCR analyses with primers flanking intron 1 of POLRMT in human cell lines. Even when using the same primers as in the previous study⁵, we could not detect any alternatively spliced transcript (Appendix Fig. 8).

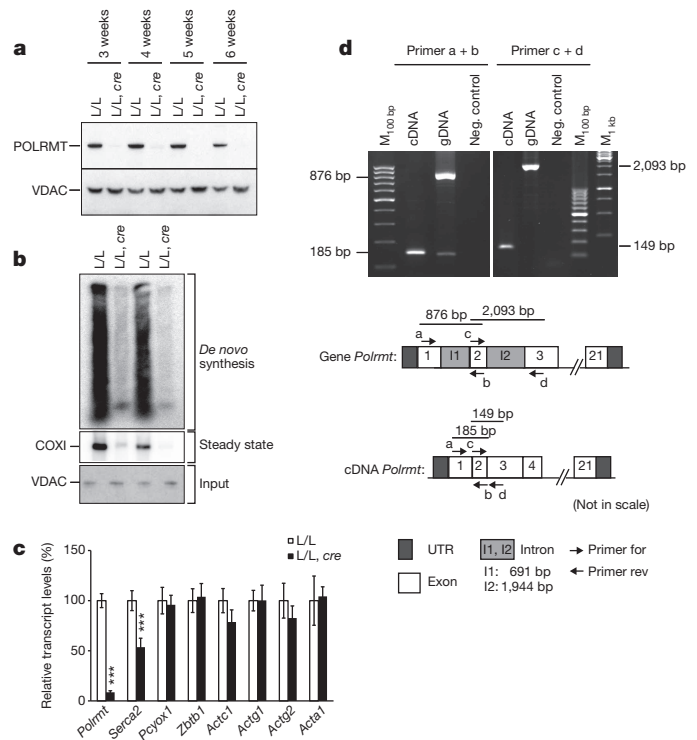


Figure 1 | Characterization of POLRMT expression. **a**, Western blot analysis of POLRMT levels in mitochondria of control (L/L) and tissue-specific knockout (L/L, cre) mouse hearts. Loading: VDAC. **b**, *De novo* transcription in mitochondria from 4-week-old control and knockout mouse hearts. Loading: VDAC. **c**, qRT-PCR quantification of nuclear transcript levels in control and knockout mouse hearts. Normalization: β -2-microglobulin (B2M); error bars indicate \pm s.e.m. (***) $P < 0.001$; $n = 6$; Student's *t*-test). **d**, RT-PCR analyses of *Polrmt* transcripts from control mouse hearts. An amplification of wild-type genomic DNA is also shown.

In this study, we have disrupted *Polrmt* and report abolished *de novo* transcription of mtDNA, thus demonstrating that there is no other intramitochondrial RNA polymerase that can replace POLRMT. Previous reports have suggested that *Polrmt* also encodes spRNAP-IV, an isoform with an essential role in transcription of important nuclear genes^{5–7}. Contrary to these previous studies, we report that neither the previously reported splice variant of the *Polrmt* gene nor the POLRMT protein isoform of ~110 kDa is detectable in mouse and human cells. Consistent with these findings, disruption of *Polrmt* has no effect on transcription of the nuclear genes that previously have been reported to be transcribed by spRNAP-IV. We therefore conclude that *Polrmt* only encodes POLRMT, which is exclusively localized to mitochondria where it is essential for mtDNA transcription^{9–11}.

Methods

To generate conditional *Polrmt* knockout mice, exon 3 was loxP-flanked (Taonic Artemis GmbH, Cologne, Germany) and excised by breeding to mice expressing cre from the muscle creatinine kinase promoter¹². Characterization of mice^{12–14} and subcellular fractionation¹⁵ were as previously described. Antibodies against mouse POLRMT (raised against recombinant mouse POLRMT and affinity purified), human POLRMT (Abcam), VDAC (Calbiochem), tubulin and histone H3 (Sigma) were used for western blots. Fixed HeLa cells were incubated with antibodies against POLRMT

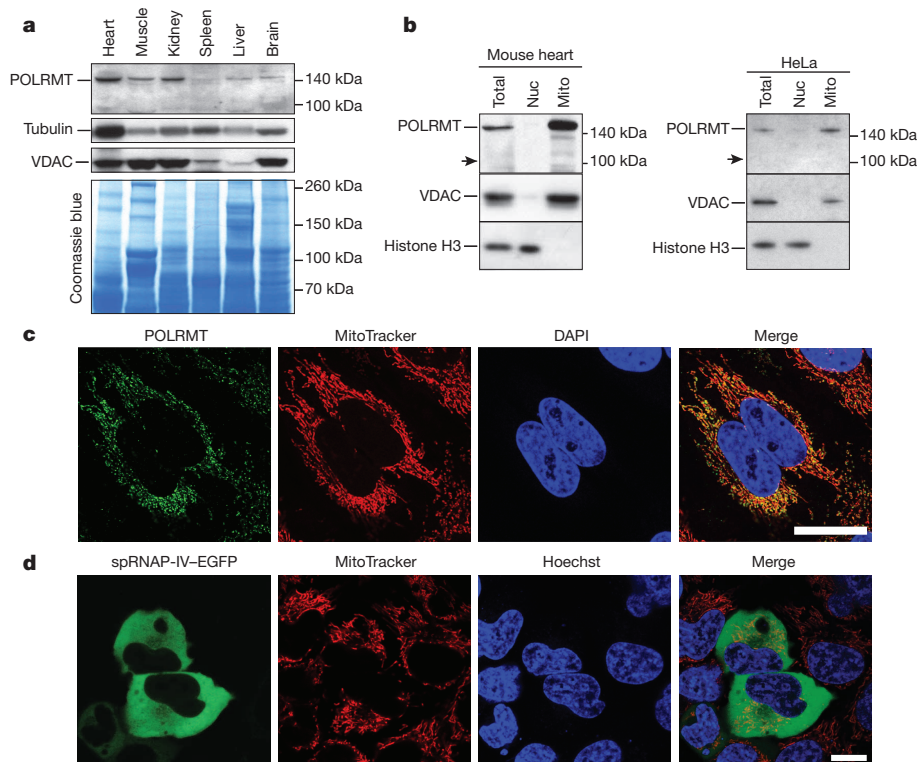


Figure 2 | Subcellular localization of POLRMT. **a**, Western blot analysis of POLRMT levels in total protein extracts from different tissues of 40-week-old wild-type mice. Loading: VDAC, tubulin and Coomassie blue. **b**, Western blot analysis of subcellular fractions from 4-week-old control mouse hearts and HeLa cells. Loading and purity of fractions: histone H3 and VDAC; Nuc,

nuclear; Mito, mitochondrial fraction. Arrow indicates ~110 kDa.

c, Immunostaining with antibodies to detect endogenous expression of POLRMT in HeLa cells. Scale bar, 25 μ m. **d**, Live-cell imaging of HeLa cells expressing spRNP-IV fused to EGFP. Scale bar, 10 μ m.

under different fixation conditions (Appendix Fig. 9). HeLa cells transfected with a construct expressing human spRNP-IV fused to EGFP were imaged with a TCS SP5-X microscope (Leica Microsystems).

Inge Kühl¹, Christian Kukat¹, Benedetta Ruzzenente¹,
Dusanka Milenkovic¹, Arnaud Mourier¹, Maria Miranda¹,
Camilla Koolmeister², Maria Falkenberg^{1,3} & Nils-Göran Larsson^{1,2}

¹Department of Mitochondrial Biology, Max Planck Institute for Biology of Ageing, 50931 Cologne, Germany.

email: larsson@age.mpg.de

²Department of Laboratory Medicine, Karolinska Institutet, 17177 Stockholm, Sweden.

³Department of Medical Biochemistry and Cell Biology, Göteborgs Universitet, 40530 Göteborg, Sweden.

Received 5 November 2013; accepted 27 June 2014.

1. Falkenberg, M., Larsson, N.-G. & Gustafsson, C. M. DNA replication and transcription in mammalian mitochondria. *Annu. Rev. Biochem.* **76**, 679–699 (2007).
2. Park, C. B. & Larsson, N.-G. Mitochondrial DNA mutations in disease and aging. *J. Cell Biol.* **193**, 809–818 (2011).
3. Liu, Z. & Butow, R. A. Mitochondrial retrograde signaling. *Annu. Rev. Genet.* **40**, 159–185 (2006).
4. Ryan, M. T. & Hoogenraad, N. J. Mitochondrial-nuclear communications. *Annu. Rev. Biochem.* **76**, 701–722 (2007).
5. Kravchenko, J. E., Rogozin, I. B., Koonin, E. V. & Chumakov, P. M. Transcription of mammalian messenger RNAs by a nuclear RNA polymerase of mitochondrial origin. *Nature* **436**, 735–739 (2005).

6. Lee, Y.-L., Chiao, C.-H. & Hsu, M.-T. Transcription of muscle actin genes by a nuclear form of mitochondrial RNA polymerase. *PLoS ONE* **6**, e22583 (2011).
7. Kravchenko, J. E. & Chumakov, P. M. Alternative transcripts of *POLRMT* gene coding for nuclear RNA polymerase IV. *Mol. Biol. (Mosk.)* **39**, 67–71 (2005).
8. Arai, M., Matsui, H. & Periasamy, M. Sarcoplasmic reticulum gene expression in cardiac hypertrophy and heart failure. *Circ. Res.* **74**, 555–564 (1994).
9. Ringel, R. *et al.* Structure of human mitochondrial RNA polymerase. *Nature* **478**, 269–273 (2011).
10. Shi, Y. *et al.* Mammalian transcription factor A is a core component of the mitochondrial transcription machinery. *Proc. Natl Acad. Sci. USA* **109**, 16510–16515 (2012).
11. Schwinghammer, K. *et al.* Structure of human mitochondrial RNA polymerase elongation complex. *Nature Struct. Mol. Biol.* **20**, 1298–1303 (2013).
12. Milenkovic, D. *et al.* TWINKLE is an essential mitochondrial helicase required for synthesis of nascent D-loop strands and complete mtDNA replication. *Hum. Mol. Genet.* **22**, 1983–1993 (2013).
13. Metodiev, M. D. *et al.* Methylation of 12S rRNA is necessary for *in vivo* stability of the small subunit of the mammalian mitochondrial ribosome. *Cell Metab.* **9**, 386–397 (2009).
14. Ruzzenente, B. *et al.* LRPPRC is necessary for polyadenylation and coordination of translation of mitochondrial mRNAs. *EMBO J.* **31**, 443–456 (2012).
15. Cox, B. & Emili, A. Tissue subcellular fractionation and protein extraction for use in mass-spectrometry-based proteomics. *Nature Protocols* **1**, 1872–1878 (2006).

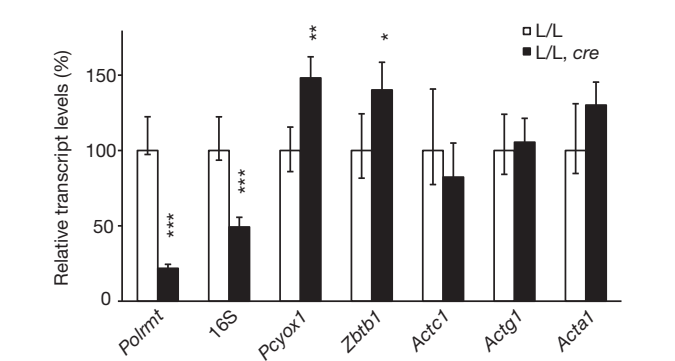
Author Contributions C.Ku. did microscopic analysis; B.R., D.M., M.M. and A.M. helped with experimental work and were involved in project planning; C.Ko. performed mouse breedings; M.F. produced the recombinant POLRMT protein used for antibody production; and I.K. performed experimental work, project planning, data analysis and wrote the manuscript together with N.-G.L.

Competing Financial Interests Declared none.

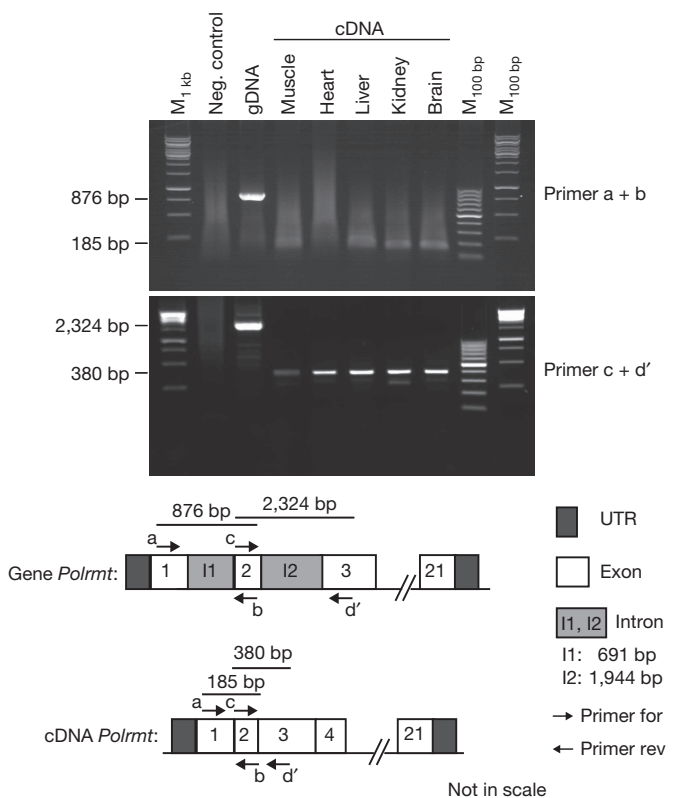
doi:10.1038/nature13690

Appendix

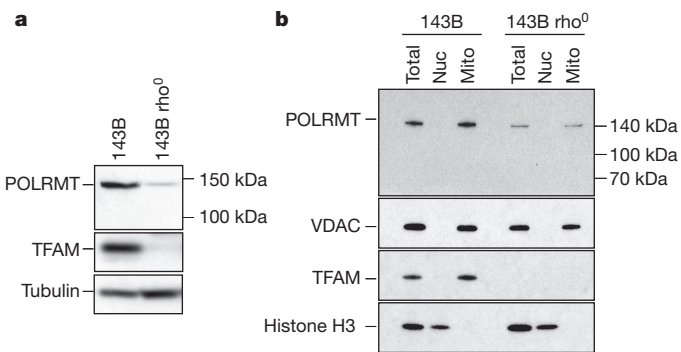
The Appendix contains Appendix Figs 1–9.



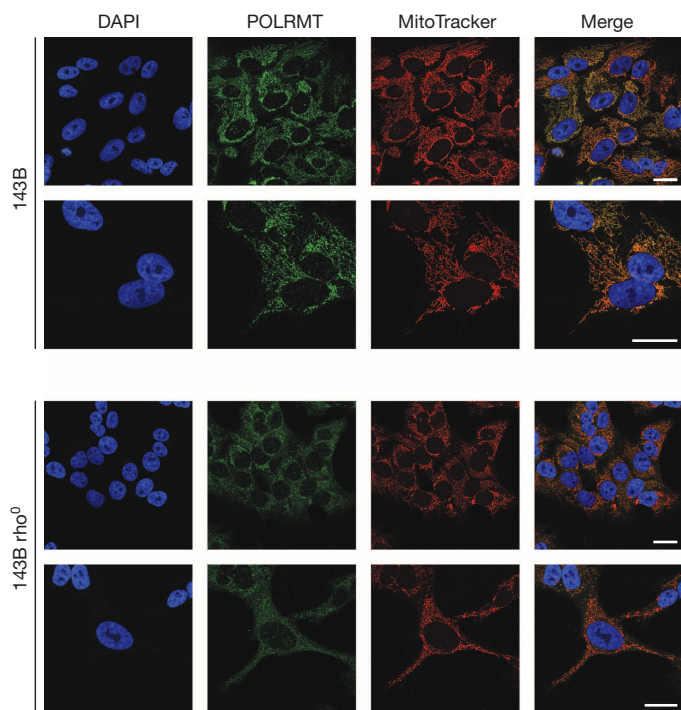
Appendix Figure 1 | Absence of POLRMT in skeletal muscle does not lead to decreased levels of transcripts from nuclear genes proposed to be transcribed by spRNAP-IV. Levels of nuclear transcripts were assessed by qRT-PCR analysis of RNA from skeletal muscle of control (L/L) and conditional knockout (L/L, *cre*) mice. The 16S rRNA is encoded by mtDNA whereas the remaining transcripts are nucleus encoded. Normalization: β -2-microglobulin (B2M); error bars indicate \pm s.e.m. (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$; $n = 6$; Student's *t*-test).



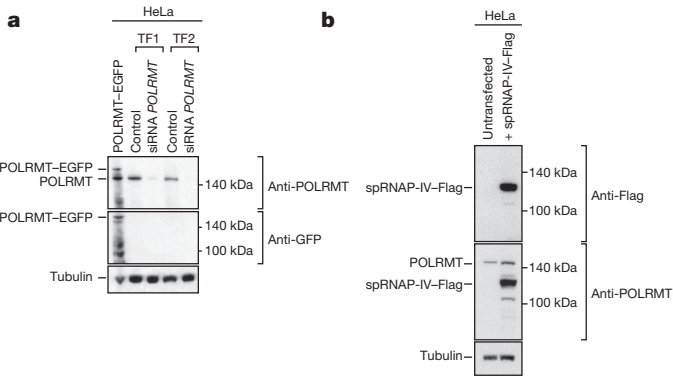
Appendix Figure 2 | Absence of transcript isoforms that could encode spRNAP-IV in different mouse tissues. RNA (cDNA) and genomic DNA (gDNA) from different mouse tissues were analysed by PCR with different sets of primers. Primers used: a, 5'-CTTGCCCGGGTCTGCGCTCC-3'; b, 5'-TCCAGCAGTTCAGCATGGCC-3'; c, 5'-GTGGCTTCTGCAGCTCAAGA-3'; d', 5'-CTTCACCCATCTCAG-3'.



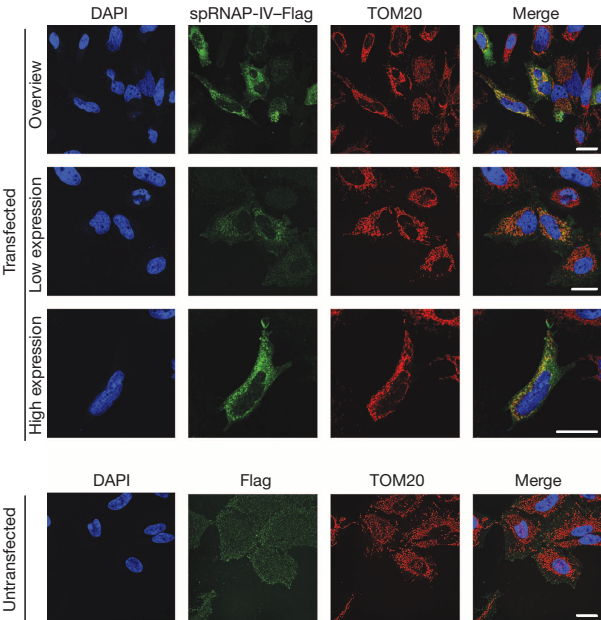
Appendix Figure 3 | Absence of the spRNAP-IV protein isoform in different human cell lines. a, Western blot analysis of total protein extracts from 143B and 143B rho0 human cells. The rho0 cells contain lower steady-state levels of POLRMT protein than wild-type cells. b, Western blot analysis of subcellular fractions (Nuc, nuclear; Mito, mitochondrial fraction) of 143B and 143B rho0 human cells incubated with antibodies against POLRMT and TFAM. TFAM is as expected absent in rho0 cells due to loss of mtDNA. Loading and purity of fractions: VDAC and histone H3.



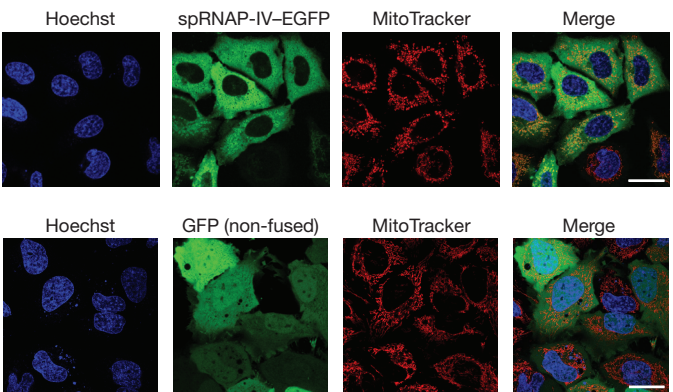
Appendix Figure 4 | Analysis of POLRMT localization in human rho0 cells. Immunostaining to detect endogenous expression of POLRMT in 143B and 143B rho0 human cells. The rho0 cells contain lower steady-state levels of POLRMT protein than wild-type cells. The cells were stained with MitoTracker red CMXRos, fixed and incubated with POLRMT antibody. The primary antibody was visualized with Alexa Fluor 488 goat anti-rabbit secondary antibody. Cells were also stained with DAPI. Scale bars, 25 μ m.



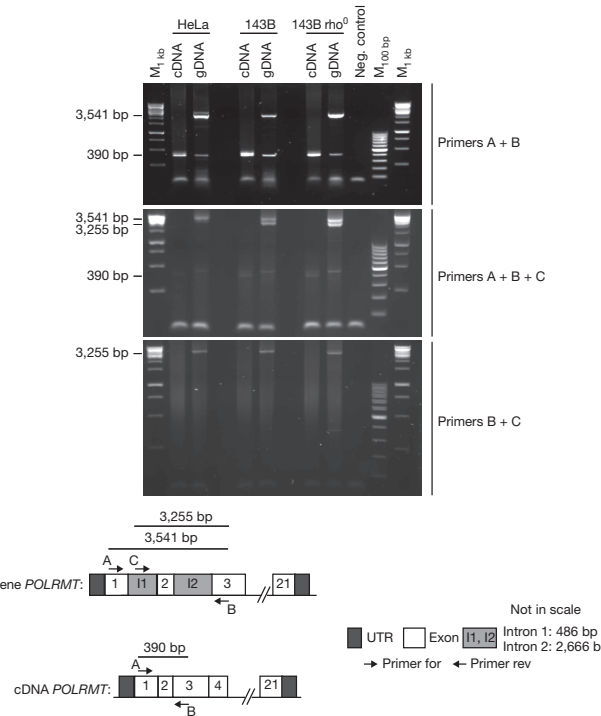
Appendix Figure 5 | The human POLRMT antibody detects both POLRMT and spRNAP-IV-Flag in human cells. **a**, Western blot analysis of total protein extracts isolated from HeLa cells transfected once (TF1) or twice (TF2) with short interfering (si)RNA against human *POLRMT* or a control siRNA. In addition, cells were transfected with a plasmid encoding human POLRMT-EGFP (left lane). The POLRMT-EGFP fusion protein is ~28 kDa larger than the endogenous POLRMT protein. **b**, Western blot analysis of HeLa cells transfected with a plasmid construct encoding spRNAP-IV with a Flag tag at its C terminus. This spRNAP-IV-Flag protein migrates at ~130 kDa. Two additional, less abundant, shorter proteins of lower molecular mass were detected with both anti-Flag and anti-POLRMT. These shorter proteins represent degradation products with preserved C termini. Membranes were incubated with antibodies against GFP or Flag M2 and human POLRMT. Loading: tubulin.



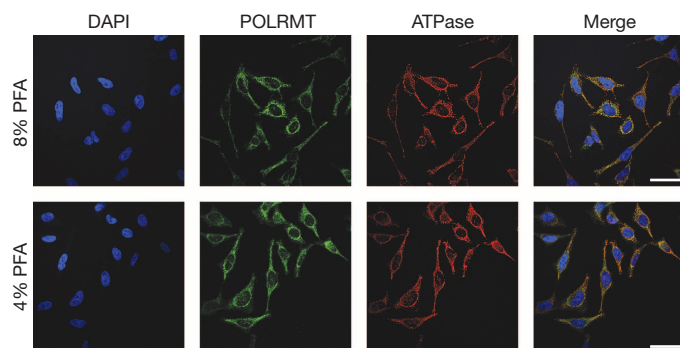
Appendix Figure 7 | Detection of spRNAP-IV-Flag expressed by transfection of human cells. HeLa cells were seeded on coverslips, transfected with spRNAP-IV-Flag, which encodes a fusion protein consisting of spRNAP-IV with an in-frame addition of the Flag peptide at its C terminus. Cells were fixed and incubated with antibodies against Flag M2 and TOM20, and stained with DAPI. Primary antibodies were visualized with Alexa Fluor 488 goat anti-mouse antibodies or Alexa Fluor 594 goat anti-rabbit antibodies. Scale bars, 25 µm.



Appendix Figure 6 | Additional exemplary cells expressing spRNAP-IV-EGFP and non-fused GFP. Live-cell imaging of HeLa cells expressing spRNAP-IV fused to EGFP (top panel) and HeLa cells expressing non-fused GFP (bottom panel). A construct expressing human spRNAP-IV was cloned into pEGFP (Clontech) to generate the spRNAP-IV-EGFP vector, which encodes a fusion protein consisting of spRNAP-IV with an in-frame addition of enhanced green fluorescent protein (EGFP) at its C terminus. The pAcGFP1-N1 plasmid (Clontech) was used to express a non-fused GFP protein. Mitochondria were counterstained with 100 nM MitoTracker Deep red FM (Lifetech). Nuclear staining was performed with Hoechst 33342 (Sigma). Live-cell imaging was performed with a Leica TCS SP5-X inverted confocal microscope (Leica Microsystems). Scale bars, 25 µm.



Appendix Figure 8 | Absence of alternative transcript isoforms encoding spRNAP-IV in human cell lines. RNA (cDNA) and genomic DNA (gDNA) from HeLa, 143B and 143B rho⁰ human cells were analysed by PCR with different sets of primers. Primers are identical to those previously used by ref. 5.



Appendix Figure 9 | Comparison of immunofluorescence results obtained after fixation with 4% or 8% formaldehyde. HeLa cells were fixed with 4% or 8% paraformaldehyde and incubated with antibodies against POLRMT or mitochondrial ATP synthase. Primary antibodies were visualized by Alexa Fluor 488 goat anti-rabbit and Alexa Fluor 594 goat anti-mouse secondary antibodies (Lifetech). Cells were also stained with DAPI. Scale bars, 50 μ m.

ARCHAEOLOGY

Art on the move

Studies of stencils and paintings from prehistoric caves in Indonesia date the art to at least 39,900 years ago — around the same age as the earliest cave art previously known, 13,000 kilometres away in western Europe. [SEE LETTER P.223](#)

WIL ROEBROEKS

The Maros karst in Sulawesi, Indonesia, is a limestone area with many caves and a large body of rock art. This art was first reported in the 1950s, and it was long assumed to be less than 10,000 years old, because it was thought that rapid erosion rates in a tropical karst environment would prevent the survival of older cave paintings. In this issue, Aubert *et al.*¹ (page 223) actually date some of that art, and report that it is one of the oldest examples of cave art in the world. This spectacular finding suggests that the making of images on cave walls was already a widely shared practice 40,000 years ago.

Mineral-rich water trickling over cave walls can form thin layers of calcite containing trace amounts of uranium. The radioactive decay of uranium atoms acts like a clock, enabling dating of the calcite formations (also called speleothems) using the uranium–thorium dating method. In cases where calcite overlies cave paintings, dating its formation can yield a minimum age for the art.

In their study, Aubert and colleagues removed tiny samples from Maros rock-art panels using a rotary tool equipped with a diamond saw blade. The coralloid speleothems (known as cave popcorn) covering the art were less than 10 millimetres thick, and the samples were subsequently micro-excavated in the lab in ‘spits’ of less than 1 mm. This method, which proceeded from the exterior surface of the speleothem towards the pigment layer and sampled above, and sometimes also below, the pigment, yielded a robust minimum and in some cases a maximum age for the paintings.

The results were unexpected. One stencilled hand was painted at least 39,900 years ago, and images of a pig deer (babirusa) and a large, indeterminate animal, probably a pig, were created at least 35,400 and 35,700 years ago, respectively. These dates are in the age range of the earliest cave art found in the westernmost tip of Europe. There, comparable dating techniques applied to calcite overlying rock art in 11 caves in northern Spain established a minimum date of 40,800 years for a red disk from El Castillo, the oldest cave painting known so far². A hand stencil from the *Panel de los Manos* at the same site yielded a minimum date



Figure 1 | Prehistoric paintings. Aubert *et al.*¹ find that the oldest-known cave art in the Maros cave in Sulawesi, Indonesia, is of comparable age — around 40,000 years old — to the previously oldest-known paintings by humans, from El Castillo cave in Spain. It is not clear whether rock art was already part of the cultural repertoire of modern humans colonizing Eurasia from Africa, or whether artistic ability arose independently in various regions. There is extensive evidence of occupation by modern humans in Oceania from around this time and somewhat longer ago (archaeological sites with dates to 40,000 years and earlier denoted by white dots; sites mentioned in the text denoted by red dots). Lower sea levels in this period (approximately –60 metres; shaded area) meant that several present-day islands were connected, but vast stretches of open sea still had to be crossed by humans colonizing the region. (Map based on data in ref. 20.)

of 37,300 years. A study³ earlier this year claims to have identified an abstract pattern engraved by Neanderthals more than 39,000 years ago, in Gorham's Cave in Gibraltar. However, both the Neanderthal authorship (on the basis of its age) and the symbolic character of this ‘rock art’ have been questioned⁴.

The earliest figurative rock art from western Europe is a painted rhinoceros from the Chauvet Cave in France, radiocarbon dated to $32,410 \pm 720$ ¹⁴C years before present⁵, which is 35,300 to 38,827 calendar years ago² (although doubts have been raised⁶ about whether the art is in fact this old). Rock fragments with traces

of red paint from the Italian site of Fumane indicate that paintings were produced there between 36,000 and 41,000 years ago⁷.

A rich corpus of rock art also exists at the other side of the spatial distribution of modern humans, in Australia. But although the first human occupation there goes back about 50,000 years, no rock art older than 30,000 years is known. Nevertheless, worn ochre crayons recovered from 50,000-year-old deposits in Arnhem Land, in northern Australia, show that some form of pigment use did occur there too, from the very first occupation onward⁸.

For the moment, the bottom line is that

cave art was practised in Europe and in southeast Asia at about the same time, before 40,000 years ago. That by itself is an important observation. But how to interpret this long-distance ‘contemporaneity’ is unclear. Southeast Asia was already occupied by the extinct hominin species *Homo erectus* at least 1 million years ago, and modern humans (members of our own species originating in Africa) reached this area probably sometime before 50,000 years ago. The modern-human occupation history of southeast Asia and the continent Sahul, which existed during periods of lower sea levels in the Pleistocene epoch (around 2.5 million to 12,000 years ago) and is now New Guinea, Australia and other islands, testifies to the role of marine navigation over vast stretches of open sea in this colonization process⁹ (Fig. 1).

Whether rock art was an integral part of the cultural repertoire of colonizing modern humans, from western Europe to southeast Asia and beyond, or whether such practices developed independently in various regions, is unknown. What is clear is that no figurative art is known from before the time of the initial expansion of *Homo sapiens* into Asia and across Europe — neither from earlier *H. sapiens* in Africa nor from their contemporaries in western Eurasia, the Neanderthals, who became extinct during the period of modern-human expansion out of Africa. The dating technique applied by Aubert *et al.* requires only minute amounts of calcite, and hence holds great potential for dating rock art worldwide, to shed light on when this art first appeared as well as on how it developed through time and space.

Aubert and colleagues’ study underlines the great cultural–historical importance of the Maros area, which is under threat from large-scale limestone mining. Their findings also stress the great relevance of Asia, and especially southeast Asia, for the study of human evolution¹⁰. The huge Asian continent is the home of recent key finds, including a series of early *Homo* individuals at Dmanisi, Georgia, dating to between 1.7 million and 1.8 million years ago¹¹, and the mysterious Denisovans — members of a *Homo* species that are known only through their genetic signature¹². The oldest-known *H. sapiens* genome was obtained from a 45,000-year-old femur, discovered at Ust-Ishi in Siberia¹³. Compared with Europe, Asia has seen little fieldwork, and new finds will keep on challenging what we think we know about human evolution. Even the evolution of the Neanderthals looks more and more like an Asian phenomenon^{14,15}, with Europe’s large number of Neanderthal remains — known from a long history of intensive research — possibly reflecting repeated colonizations from central and western Asia¹⁶.

Southeast Asia also harbours the type site of *H. erectus*, at Trinil, Java, and Sulawesi’s neighbouring island, Borneo, contains the

spectacular Greater Niah Cave, which has a record of human presence from about 50,000 years ago onward, including the first unambiguous fossil of a modern human in the area, a skull at least 40,000 years old¹⁷. Borneo also has a rich, but as-yet-undated, rock-art record, with some very striking similarities to the Maros paintings¹⁸. Finally, one of the most remarkable and least expected palaeoanthropological discoveries was made 400 kilometres south of the Maros area, on the island of Flores: here, the late Mike Morwood, one of the authors of the Maros cave-art study, discovered the skeleton of a puzzling diminutive hominin, presented 10 years ago as *Homo floresiensis*¹⁹ and nicknamed ‘the hobbit’ — another illustration of the surprises that this region can offer. ■

Wil Roebroeks is in the Human Origins group, Faculty of Archaeology, Leiden University, 2300 RA Leiden, the Netherlands. e-mail: w.roebroeks@arch.leidenuniv.nl

1. Aubert, M. *et al.* *Nature* **514**, 223–227 (2014).
2. Pike, A. W. G. *et al.* *Science* **336**, 1409–1413 (2012).

3. Rodríguez-Vidal, J. *et al.* *Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1411529111> (2014).
4. Callaway, E. *Nature* <http://dx.doi.org/10.1038/nature.2014.15805> (2014).
5. Valladas, H. *et al.* *Nature* **413**, 479 (2001).
6. Pettitt, P. B. *J. Hum. Evol.* **55**, 908–917 (2008).
7. Broglio, A. *et al.* *L’Anthropologie*, **113**, 753–761 (2009).
8. David, B. *et al.* *J. Archaeol. Sci.* **40**, 3–10 (2012).
9. O’Connell, J. F., Allen, J. & Hawkes, K. in *The Global Origins and Development of Seafaring* (eds Anderson, A., Barrett, J. & Boyle, K.) 57–68 (Cambridge Univ. Press, 2010).
10. Dennell, R. W. & Porr, M. (eds) *Southern Asia, Australia and the Search for Human Origins* (Cambridge Univ. Press, 2014).
11. Lordkipanidze, D. *et al.* *Science* **342**, 326–331 (2013).
12. Reich, D. *et al.* *Nature* **468**, 1053–1060 (2010).
13. Gibbons, A. *Science* **343**, 1417 (2014).
14. Dalén, L. *et al.* *Mol. Biol. Evol.* **29**, 1893–1897 (2012).
15. Prüfer, K. *et al.* *Nature* **505**, 43–49 (2014).
16. Hawks, J. *Annu. Rev. Anthropol.* **42**, 433–449 (2013).
17. Barker, G. (ed.) *Rainforest Foraging and Farming in Island Southeast Asia: The Archaeology of the Niah Caves, Sarawak* (McDonald Inst. Archaeol. Res., 2013).
18. Fage, L.-H. & Chazine, J.-M. *Borneo: Memory of the Caves* (Le Kalimanthrope, 2010).
19. Brown, P. *et al.* *Nature* **431**, 1055–1061 (2004).
20. O’Connell, J. F. & Allen, J. *Aust. Archaeol.* **74**, 5–31 (2012).

ULTRALUMINOUS X-RAY SOURCES

Small field with a large impact

The nature of ultraluminous X-ray astronomical sources has long been unclear. The latest observations of these rare systems provide some crucial clues, but still leave theorists scratching their heads. SEE LETTERS P.198 & P.202

JEANETTE C. GLADSTONE

In the late 1970s, astronomers discovered objects that emit unusually bright X-rays¹. Given their extreme X-ray luminosity, these ultraluminous X-ray sources were thought to contain black holes. However, the mass of the black holes powering such objects has been a topic of much debate. Two studies in this issue, by Motch *et al.*² (page 198) and Bachetti *et al.*³ (page 202), together with two recent reports by Pasham *et al.*⁴ and Liu *et al.*⁵, are changing our views about these systems.

Most black holes are created during the violent deaths of massive stars. Although such stellar-mass black holes weigh about 3 to 100 times the mass of our Sun, they can be difficult to see. Their extreme gravitational pull attracts anything that strays too close, even light. So, to learn more about them, we must observe them indirectly, by studying the effect they have on their environment.

If the stellar-mass black hole is orbited by a companion star, we can study its effects on the star. The black hole can pull material from

the star’s wind and/or surface. As material falls in (accretes), forming an accretion disk, some of the material’s gravitational potential energy is lost as light — mainly X-rays. Such X-ray binary systems (Fig. 1) contain not just a disk but also an optically thin (transparent) medium, which is thought to sit either above and below the disk (a corona) or between the disk and the black hole (a hot inner flow). As the rate at which material travels through the accretion disk changes, the geometry of the system, and so its accretion state, will change accordingly.

X-ray binaries can also contain neutron stars — the smaller cousins of stellar-mass black holes. Like stellar-mass black holes, neutron stars are born in violent star deaths, but they are lighter, weighing only around 1.4 solar masses. The gravitational pull of these systems is again very strong, drawing in material. But unlike black holes, light can escape from them, and we can see their surface.

Black holes also have much heavier cousins, which reside in the centres of galaxies. They are known as supermassive black holes, and weigh

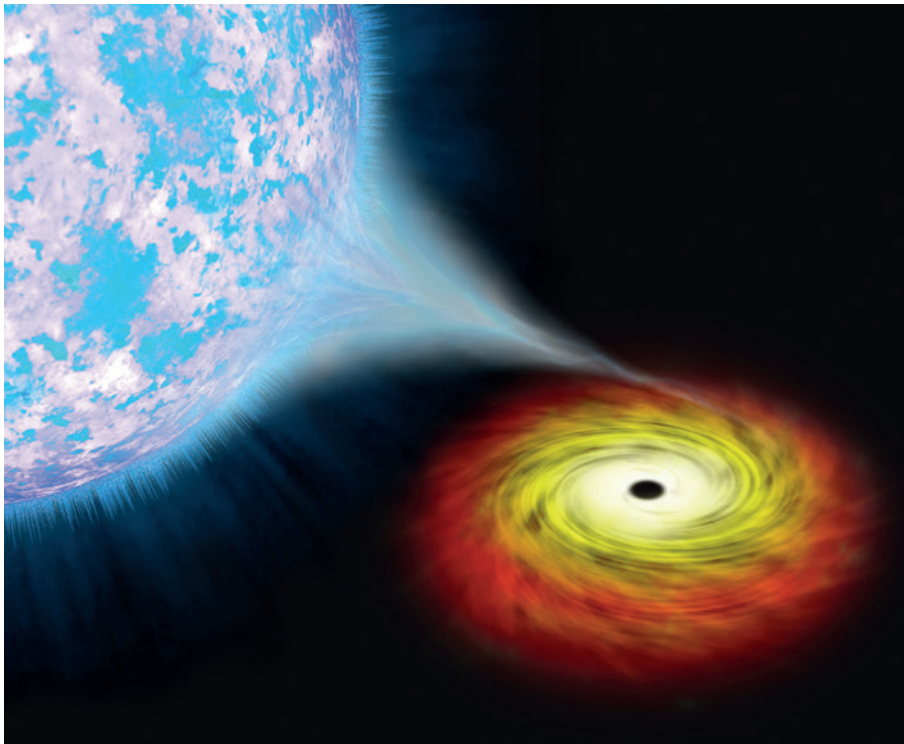


Figure 1 | Artist's impression of an ultraluminous X-ray source. Astronomical objects that emit extremely luminous X-rays, such as the objects studied by Motch *et al.*² and Bachetti *et al.*³, are thought to contain a black hole or a neutron star (right) that is drawing in material from a companion star (left). The infalling material forms an accretion disk, and loses some of its gravitational potential energy as light.

millions to billions of solar masses. They show similar accretion states to stellar-mass black holes, with their luminosity apparently scaling with their mass. Although we understand the formation of stellar-mass black holes, that of supermassive black holes is not understood.

Ultraluminous X-ray sources (ULXs) are intermediate in luminosity between stellar-mass and supermassive black holes, and so it was thought that they might be X-ray binaries containing intermediate-mass black holes (weighing hundreds to thousands of solar masses). In the past few years, evidence has been found suggesting that the shape of the X-ray spectra of these sources is distinctly different from that of previously observed accretion states^{6,7}. This suggests that we may be observing stellar-mass black holes that are behaving oddly, possibly residing in a new extreme (ultraluminous) accretion state.

Motch *et al.*² present a multi-wavelength analysis of an X-ray source dubbed P13 — a ULX in the spiral galaxy NGC 7793 that has been observed to vary dramatically in luminosity (by a factor of 40). The authors' X-ray analysis of P13 shows evidence of the ultraluminous state⁶ at its brightest X-ray emission. Furthermore, their optical data indicate that the source has a massive, luminous companion star, known as a supergiant B star. By making many optical observations over 8 years, the authors find that the star is being continually bombarded with X-ray radiation from the black hole's accretion disk, leading to X-ray

heating of the star's surface even when the X-ray luminosity is decreased. This suggests that the star is exposed to bright X-ray emission from the inner parts of the disk, even though the geometry of the system is shielding this X-ray-emitting region of the disk from Earth's line of sight. On the basis of these observations, the researchers conclude that P13 contains a black hole of around 15 solar masses, confirming the idea that ULXs can be powered by stellar-mass black holes in an extreme accretion state.

This hypothesis is supported by Liu and colleagues' study⁵ of a source called M101 ULX-1. Their optical data indicate that this source can be explained by a black hole of about 20–30 solar masses fed by a Wolf–Rayet star, an ageing massive star with strong winds. These results, together with those of Motch and colleagues², suggest that the mysterious nature of ULXs may now have been solved.

However, things are also becoming complicated. Analyses^{3,4} of two of the ULXs in the cigar-shaped galaxy M82 — M82 X-1 and M82 X-2 — show that some of these sources may be more massive or more extreme than stellar-mass black holes in an ultraluminous state.

M82 X-1 is the brightest ULX in this galaxy. Pasham *et al.*⁴ performed an X-ray study of this ULX and found two almost-periodic variations. Such quasi-periodic oscillations have previously been observed in stellar-mass black-hole systems, and seem to scale with mass⁸. Pasham and colleagues used this

mass-scaling relationship to infer a black-hole mass of about 400 times that of the Sun. This places the object in the range of intermediate-mass black holes.

Conversely, M82 X-2 seems to be more extreme than was previously thought possible. Using several X-ray telescopes, Bachetti *et al.*³ find that M82 X-2 does not contain a black hole, but instead contains a neutron star. This object was seen to be pulsing, something not possible for black holes. Such pulsing can come only from objects known as pulsars — magnetic neutron stars with fields so strong that material from the accretion disk is funnelled onto their poles, creating a hotspot. As these pulsars spin, beams of light from the hotspots sweep in and out of Earth's line of sight. Astronomers see this phenomenon as pulses of light in the same way as pulses are observed from a lighthouse as its lamp rotates. However, this system seems to be pumping out massive amounts of energy, suggesting an accretion rate that is 10 times larger than anything previously measured for these sources. It is also 100 times brighter than the original theoretical limit for such sources. This finding challenges our current theories of accretion physics, specifically of accretion onto magnetic neutron stars.

These results^{3,4} will surprise the ULX community. They show that ULXs are heterogeneous, and that more-detailed studies of individual sources can push the boundaries of our understanding, not only of the field of ULXs but also of other fields. The confirmation that stellar-mass black holes can accrete at extreme rates^{2,3} could help to unlock the mystery of the rapid growth of massive black holes in the early Universe. With the discovery of intermediate-mass black holes⁴, models of the formation of supermassive black holes and hence galaxy formation can be tested: research indicates that these two processes are linked. The finding that neutron stars can reach such high X-ray luminosities³ will probably leave theorists scratching their heads to find ways of pushing neutron-star and accretion physics to new extremes. This is an exciting time for the study of ULXs — a small field that can have a far-reaching impact. ■

Jeanette C. Gladstone is in the Department of Physics, University of Alberta, Edmonton, Alberta T6G 2E1, Canada.
e-mail: j.c.gladstone@ualberta.ca

1. Fabbiano, G. *Annu. Rev. Astron. Astrophys.* **27**, 87–138 (1989).
2. Motch, C., Pakull, M. W., Soria, R., Grisé, F. & Pietrzyński, G. *Nature* **514**, 198–201 (2014).
3. Bachetti, M. *et al. Nature* **514**, 202–204 (2014).
4. Pasham, D. R., Strohmayer, T. E. & Mushotzky, R. F. *Nature* **513**, 74–76 (2014).
5. Liu, J.-F., Bregman, J. N., Bai, Y., Justham, S. & Crowther, P. *Nature* **503**, 500–503 (2013).
6. Gladstone, J. C., Roberts, T. P. & Done, C. *Mon. Not. R. Astron. Soc.* **397**, 1836–1851 (2009).
7. Bachetti, M. *et al. Astrophys. J.* **778**, 163 (2013).
8. McHardy, I. M., Koerding, E., Knigge, C., Uttley, P. & Fender, R. P. *Nature* **444**, 730–732 (2006).

STRUCTURAL BIOLOGY

Lariat lessons

The spliceosome enzyme complex removes intron sequences from RNA transcripts to form messenger RNA. The crystal structure of a lasso-shaped RNA suggests a mechanism for this splicing process. [SEE ARTICLE P.193](#)

ROBERT T. BATEY

One of the seminal discoveries that propelled biologists towards an increased appreciation of RNA's many roles in cellular function was that genes can be interspersed with non-protein-coding sequences called introns^{1,2}. During the transcription of genomic DNA into messenger RNA, the RNA sequences generated from introns are removed by the spliceosome (a complex enzyme formed from proteins and RNA), the structure of which is poorly understood. On page 193 of this issue, Robart *et al.*³ provide insight into the mechanism of this splicing process by reporting the crystal structure of a self-splicing group II intron — an RNA enzyme that shares a common ancestor with spliceosomal RNAs.

The evolutionary relationship between the spliceosome and group II introns was first inferred from their use of the same two-step chemical mechanism for splicing (Fig. 1). A defining feature of this process is the creation of a 2'–5' phosphodiester linkage at an adenosine nucleotide in the intron's 'branch site' during the first step of splicing. This linkage causes the excised intron to adopt a characteristic lariat structure — a loop with a branch attached at the branch site.

If the splicing reaction was the only reaction performed by group II introns, the lariat intermediate would not be required. Indeed, an evolutionarily primitive class of group II intron yields a linear splicing product, rather than a loop⁴. But a key advantage of the lariat is that it makes the first step of the splicing reaction reversible, allowing group II introns to invade DNA through a reverse splicing mechanism and so disperse throughout the genome. This mobility of group II introns is thought to have dramatically shaped the early evolution of eukaryotes⁵ (organisms that include fungi, plants and animals), as demonstrated, in part, by the fact that their descendants comprise at least half of the human genome⁶.

Group II introns are defined by a common secondary structure divided into six functional domains (Fig. 2). The core of these RNA molecules is formed principally by domains I and V. Domain I acts as a scaffold to organize the other domains and exons (the protein-coding sequences attached at either end of the intron) around the catalytic site housed in domain V. The heart of group II introns is a set of three

evolutionarily conserved base triples (structures analogous to base pairs, but involving three bases), which host a set of metal cations that directly participate in catalysis of the splicing reactions (a 'catalytic triplex'; Fig. 2). The structural details of the hydrolytic splicing mechanism were previously revealed through analysis of the crystal structure of a group II intron from the bacterium *Oceanobacillus iheyensis*⁷. A biochemical analysis⁸ published this year revealed that the active site of the spliceosome is similar to that of the *O. iheyensis* intron, supporting their evolutionary relationship.

Robart and co-workers have determined the atomic-level structure of a group II intron from the brown alga *Pylaiella littoralis*, to understand the lariat's role in splicing. This RNA is a member of one of the predominant classes of group II intron that mainly use a branch-site

adenosine in domain VI for splicing, and it is thus more representative of the structure and mechanism of these RNA enzymes than the *O. iheyensis* intron. The structure depicts the product of the splicing reaction, in which the lariat form of the intron is bound to the spliced exons through base-pairing interactions with domain I. This intron contains domains that are highly truncated (domains II and III) or entirely absent (domain VI) in the *O. iheyensis* structure, which means that it provides a richer understanding of how these RNAs achieve fully reversible splicing with high fidelity.

The authors reveal how peripheral extensions not present in the *O. iheyensis* intron play a significant part in supporting catalysis. Domains II and III have been implicated in enhancing the catalytic efficiency of the intron through interactions with domains VI and V, respectively⁹. Robart and colleagues' structure strikingly reveals that domain II — whose function was poorly understood — serves as a nexus for a set of four 'tetraloop-mediated' interactions with domains I, III and VI. These types of interaction help to establish the higher-order architecture of many biological RNAs by packing helices together. In the present case, such interactions seem to substantially buttress the RNA architecture around the intron's active site.

The most important finding relates to how

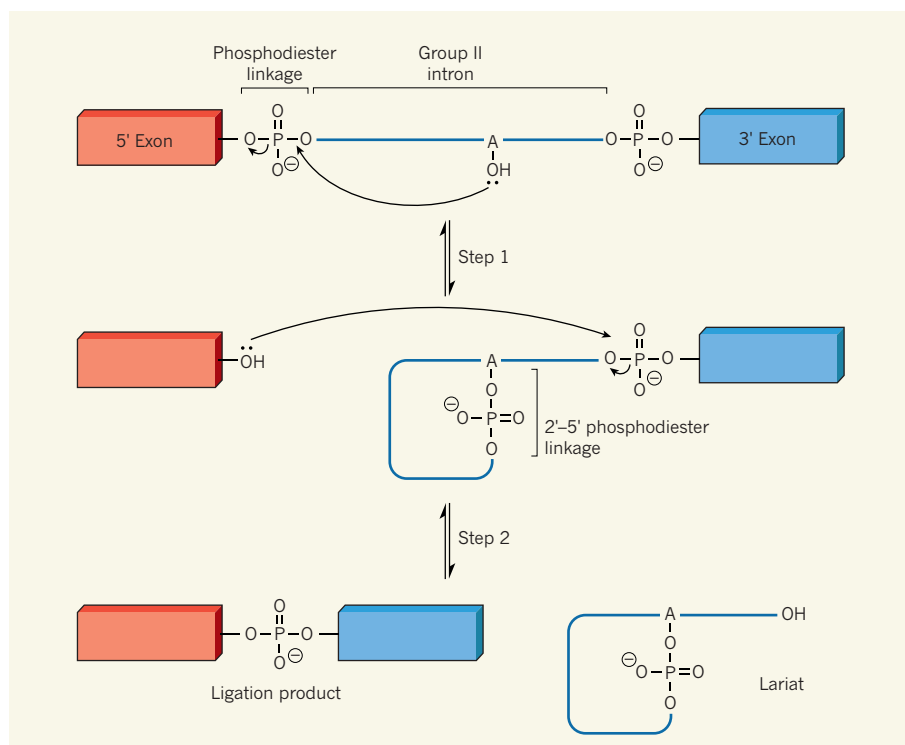


Figure 1 | Self-splicing reactions of group II introns. RNA enzymes called group II introns catalyse their own removal from RNA and join together flanking exon sequences. In the first step of this splicing process, the hydroxyl group (OH) of an adenosine nucleotide (A) at the branch site of the intron attacks the phosphodiester linkage at the 5' splice site. This produces a free 5' exon and a structure in which the 5' end of the intron is linked to the attacking adenosine through a newly formed 2'–5' phosphodiester linkage. In the second step, the 3' hydroxyl group of the 5' exon attacks the phosphodiester linkage at the 3' splice site to yield the ligation product (in which the exons are joined together) and the lariat form of the intron. Curly arrows indicate electron movement; dots indicate lone pairs of electrons.

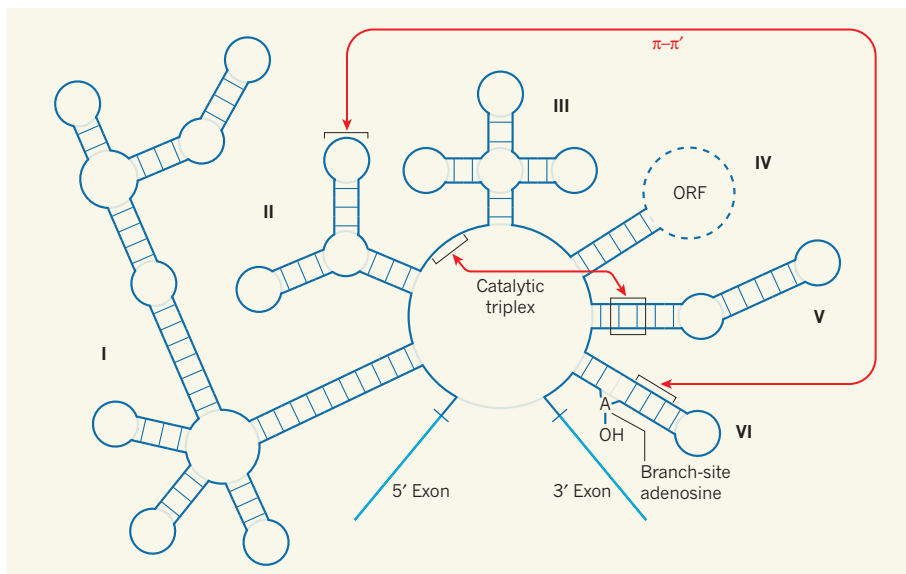


Figure 2 | Cartoon of the domain organization of group II introns. The domains are numbered I to VI. A long-range interaction between three highly conserved base pairs in domain V and nucleotides in the single-stranded region between domains II and III forms part of a structure called a catalytic triplex, which is the RNA core of the active site. Robart *et al.*³ identify an interaction (the π - π' interaction) between a loop in domain II and a section of domain VI that is adjacent to the branch-site adenosine, and propose that this interaction causes the intron to switch between the two splicing steps. Domain IV often carries a gene (ORF; broken line) encoding a protein that facilitates efficient catalysis of splicing or mobility of the group II intron. Short lines between folded sections of the intron indicate base-pairing; more base pairs form in the structure than are shown here.

the branch-site adenosine promotes reversible splicing. The authors dub the newly defined tetraloop-mediated interaction between domains II and VI the π - π' interaction (Fig. 2). The main job of domain VI is to position the branch-site adenosine within the catalytic core for its attack on the 5' splice site⁹. This domain is analogous to a helix formed in the spliceosome between one of the spliceosome's constituent RNAs and an intron around the intron's branch-site adenosine¹⁰. When Robart *et al.* simultaneously disrupted the π - π' interaction and a second tetraloop-mediated interaction in their introns, they observed almost full blockade of the second splicing step.

This observation, along with the authors' low-resolution structure of the intron in the pre-catalytic state, reveals that the π - π' interaction is the key structural switch in the transition from the first to the second step of splicing. The researchers propose that formation of the branched adenosine directly promotes the π - π' interaction, causing the branch site to move out of the catalytic site to a position about 20 Å away, as observed in the crystal structure of the intron's product form. This pulls the 3' splice site into the active site, positioning the 5' exon ready for the second splicing step. However, the precise mechanism by which lariat formation remodels the active site remains a question for future studies. A set of high-resolution structures encompassing each step in the splicing reaction will be required to reveal the full details of lariat-mediated splicing.

So what does the lariat structure teach us

about splicing by the spliceosome? There is no known interaction in spliceosomal RNAs that is analogous to the π - π' interaction, begging the question of how formation of the branch site in the spliceosome triggers switching between the two steps. Like many modern RNAs, those in the spliceosome have evolved to function in concert with accessory proteins¹¹. Although RNA is almost certainly at the heart of the spliceosome, these proteins clearly have crucial roles in the regulation and specificity of splicing, and in facilitating the formation of RNA architecture. It is distinctly

possible that one of the proteins next to the active site drives this movement.

One potential candidate for this is Prp8, a protein found next to the spliceosome's active site and whose structure has been determined¹². Unfortunately, there is insufficient biochemical and genetic information to indicate how Prp8 interacts with the RNA at the active site of the spliceosome. However, advances made in cryo-electron microscopy¹³ should put the structure of the spliceosome within reach. In the meantime, Robart and colleagues' structure of the lariat form of a group II intron provides tantalizing clues about one of the most complex macromolecular machines in biology. ■

Robert T. Batey is in the Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado 80309-0596, USA.
e-mail: robert.batey@colorado.edu

- Berget, S. M., Moore, C. & Sharp, P. A. *Proc. Natl Acad. Sci. USA* **74**, 3171–3175 (1977).
- Chow, L. T., Gelinas, R. E., Broker, T. R. & Roberts R. J. *Cell* **12**, 1–8 (1977).
- Robart, A. R., Chan, R. T., Peters, J. K., Rajashankar, K. R. & Toor, N. *Nature* **514**, 193–197 (2014).
- Podar, M., Chu, V. T., Pyle, A. M. & Perlman, P. S. *Nature* **391**, 915–918 (1998).
- Martin, W. & Koonin, E. V. *Nature* **440**, 41–45 (2006).
- Lambowitz, A. M. & Zimmerly, S. *Cold Spring Harb. Perspect. Biol.* **3**, a003616 (2011).
- Toor, N., Keating, K. S., Taylor, S. D. & Pyle, A. M. *Science* **320**, 77–82 (2008).
- Fica, S. M., Mefford, M. A., Piccirilli, J. A. & Staley, J. P. *Nature Struct. Mol. Biol.* **21**, 464–471 (2014).
- Pyle, A. M. *Crit. Rev. Biochem. Mol. Biol.* **45**, 215–232 (2010).
- Query, C. C., Moore, M. J. & Sharp, P. A. *Genes Dev.* **8**, 587–597 (1994).
- Cech, T. R. *Cold Spring Harb. Symp. Quant. Biol.* **74**, 11–16 (2009).
- Galej, W. P., Oubridge, C., Newman, A. J. & Nagai, K. *Nature* **493**, 638–643 (2013).
- Amunts, A. *et al. Science* **343**, 1485–1489 (2014).

This article was published online on 24 September 2014.

PARTICLE PHYSICS

The mass of a top

A measurement of the mass of the heaviest-known elementary particle, the top quark, which exists for less than a trillionth of a trillionth of a second, sheds light on the ultimate fate of our Universe, although ambiguities cloud its interpretation.

PETER SKANDS

Writing in *Physical Review Letters*, researchers working in the D0 experiment (Abazov *et al.*¹) at the Tevatron accelerator at Fermilab near Chicago, Illinois, report the most precise single measurement so far of the mass of the heaviest-known elementary particle, the top quark. The result concludes an exciting 20-year saga

— from the joint discovery of the top quark by the D0 Collaboration² and its competitor the CDF Collaboration³, to a measurement of the top quark's mass with a precision better than 0.5%. A similar result from the CDF experiment is to be expected, updating their 2012 result⁴.

The top quark is one of six types of quark predicted by the standard model of particle physics; quarks are elementary particles that

make up composite particles such as protons and neutrons. The top quark existed in the extremely hot conditions of the early Universe and can be recreated artificially by large particle accelerators such as the Tevatron. The D0 experiment takes its name from its location on the accelerator ring. According to the D0 measurement, a top quark weighs 187.85 ± 0.82 atomic units (174.98 ± 0.76 giga-electronvolts c^{-2} in particle-physics units, with c being the speed of light), just shy of the mass of a gold atom. Unlike atoms, however, the top quark is elementary, and acquires its mass by interacting with the elusive, omnipresent Higgs field, the telltale evidence of which — the Higgs boson — was famously discovered^{5,6} in 2012.

Briefly stated, the presence of the Higgs field in the Universe causes an increase in the potential energy of all particles except photons, gluons and possibly neutrinos. The extra potential energy is equivalent to a mass, and the size of this mass is proportional to the strength of the Higgs field (called its vacuum expectation value, a universal constant) and to the size of each particle's 'Higgs charge', which determines how strongly each particle interacts with the Higgs field. For the top quark, this charge is called the top-quark Yukawa coupling, named after the Japanese physicist and Nobel laureate Hideki Yukawa.

The fact that the top quark has by far the largest mass among elementary particles implies that it has by far the largest Yukawa coupling, and this in turn gives rise to some of the most significant quantum fluctuations in nature. At the quantum level, the Higgs field constantly fluctuates into pairs of 'virtual' particles and antiparticles, which are allowed a brief existence by Heisenberg's uncertainty principle. Because of the huge top-quark Yukawa coupling, the fluctuations involving the top quark affect the shape of the Higgs potential (which describes the potential energy of the Higgs field as a function of the field strength). In fact, making the strong assumption that there are no as-yet-undiscovered particles, the Higgs field seems to exist in a local minimum of the potential⁷, which would make the Universe as we know it unstable. Fortunately, the calculated lifetime of the Universe comfortably exceeds its present age. So, rest assured, the Universe will not decay tomorrow. But the desire to ascertain its ultimate fate is a key reason for accurately determining the top-quark mass: differences of just 5% in this mass make the difference between stability and instability. With modern experimental measurements such as that from D0 reaching precisions better than 1%, it is now largely a matter of improving the delicate theoretical calculations to settle the issue.

The measurement is far from trivial. First, in collisions of protons and antiprotons at the Tevatron, top quarks are created predominantly by the strong nuclear force, which conserves 'quark number', and so they must

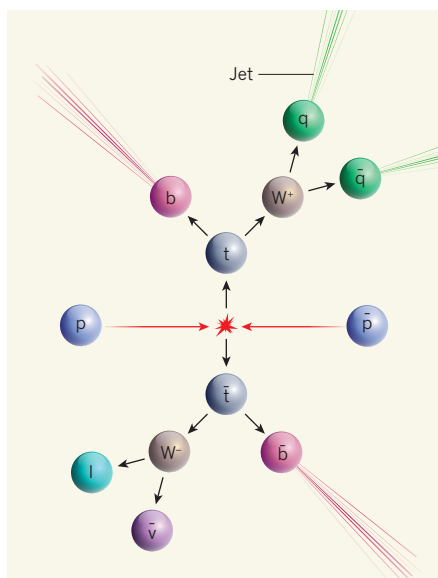


Figure 1 | Decay of the top quarks. The collisions of protons (p) and antiprotons (\bar{p}) in the Tevatron accelerator can, among many other possible reactions, produce a pair of top quarks, one top quark (t) and one antitop quark (\bar{t}), which both decay rapidly to lighter particles. By measuring the energies and momenta of these particles accurately, the masses of the original top quarks can be reconstructed. The specific decay pattern shown here corresponds to the 'leptons + jets' channel used by the D0 Collaboration measurement¹, in which the top and the antitop each decay differently. The antitop quark decays into a negatively charged W boson (W^-), which in turn decays to a charged lepton (l ; an electron, muon or tauon) and an antineutrino ($\bar{\nu}$). The top quark decays into a positively charged W boson (W^+), which decays into quark (q) and antiquark (\bar{q}) jets. Both decays also produce a bottom (b)-quark jet.

be created in pairs (top plus antitop; Fig. 1). And they decay within one-trillionth of a picosecond (1 picosecond is 10^{-12} s) into bottom quarks and W bosons, the latter being elementary particles that mediate the weak nuclear force. Bottom (b) quarks undergo a process called fragmentation and turn into jets: sprays of nuclear particles (hadrons) arranged in fractal-like patterns. D0 identifies these 'b-quark jets' with around 65% efficiency by using the fact that b-quarks travel about one centimetre before decaying, leaving a detectable 'displaced vertex'. The W bosons decay immediately, either to a charged lepton (electron, muon or tauon) accompanied by a neutrino or an antineutrino, or to a quark jet and an antiquark jet. Putting all this together, the final states are therefore complex and are classified into three main categories according to the decay products of the two W bosons: all-jets, leptons and jets, and dileptons.

The all-jets category represents half of the events, but there are significant non-top 'multi-jet' backgrounds, and accurate energy calibrations of the measured particle jets are

challenging. Dileptons can be measured very precisely, but they occur only 4% of the time, and the two escaping neutrinos (one of which is actually an antineutrino) leave an irreducible ambiguity in the determination of the top-quark mass.

The best of both worlds is obtained in the 'leptons + jets' channel (Fig. 1), which was used by the D0 Collaboration. The rate of occurrence of this channel (30%) is reasonable, there are fewer jets and lower non-top backgrounds than in the all-jets case, and momentum conservation accurately constrains the energy and direction of the single escaping neutrino.

The masses of the original top quarks are encoded in the energies and momenta of their decay products. The calibration techniques to extract the 'true' top-quark mass from the raw data have evolved enormously since 1995. In the D0 analysis, the set of measured variables is compared with state-of-the-art theoretical calculations of the likelihood of combined signal and background events for several reference top-quark mass values, to find the one that maximizes the overall likelihood.

The accuracy of the theoretical calculations therefore affects the measurement precision, and elaborate cross-checks are necessary to estimate the added uncertainty. Imperfections in the description of the jet-fragmentation process and the modelling of the energies and momenta of the top quarks and their decay products are especially important. The D0 analysis achieves an unprecedented systematic uncertainty (experimental plus theoretical) of just $0.5 \text{ GeV } c^{-2}$.

The Tevatron collider was shut down in 2011, and this measurement is the final word from D0 on the leptons + jets channel, comprising about 2,500 collision events. Meanwhile, the ATLAS and CMS experiments at the Large Hadron Collider (LHC) at CERN near Geneva, Switzerland, have already published^{8,9} measurements rivalling those of CDF⁴ and D0. When the LHC starts up again in 2015, top-quark production rates will skyrocket to thousands of times the Tevatron level.

Given the high precision being attained by these experiments, it is ironic that theorists are not quite sure how to interpret the measured values. In quantum field theory, particle masses can be defined in many slightly different ways, and currently there is no consensus about exactly which definition the experimental measurements correspond to, leaving a 1% ambiguity when converting between the measured mass and the Yukawa coupling. Because this is larger than the precision on the raw measurement, this is an issue that urgently needs to be resolved before we can truly claim that we know the mass of a top. ■

Peter Skands is at CERN, 1211 Geneva 23, Switzerland, and the School of Physics, Monash University, PO Box 27, Clayton, Victoria 3800, Australia.

e-mail: peter.skands@monash.edu

1. Abazov, V. M. *et al. Phys. Rev. Lett.* **113**, 032002 (2014).
2. Abachi, S. *et al. Phys. Rev. Lett.* **74**, 2632–2637 (1995).
3. Abe, F. *et al. Phys. Rev. Lett.* **74**, 2626–2631 (1995).

4. Aaltonen, T. *et al. Phys. Rev. Lett.* **109**, 152003 (2012).
5. ATLAS Collaboration *et al. Phys. Lett. B* **716**, 1–29 (2012).
6. CMS Collaboration *et al. Phys. Lett. B* **716**, 30–61 (2012).

7. Elias-Miró, J. *et al. Phys. Lett. B* **709**, 222–228 (2012).
8. The ATLAS Collaboration. *Eur. Phys. J. C* **72**, 2046 (2012).
9. The CMS collaboration. *J. High Energy Phys.* **105**, 1212 (2012).

HEALTH

The weighty costs of non-caloric sweeteners

Analyses in mice and humans indicate that non-caloric artificial sweeteners may promote obesity-associated metabolic changes by changing the function of the bacteria that colonize the gut. [SEE ARTICLE P.181](#)

TAYLOR FEEHLEY & CATHRYN R. NAGLER

In many parts of the world, obesity is becoming increasingly prevalent. Weight control is important for reducing the risk of metabolic diseases such as type 2 diabetes, which is characterized by high blood-glucose levels and insulin resistance. Limiting calorie intake and replacing dietary fat and sugar with low- or non-caloric alternatives is a common weight-loss strategy¹. Non-caloric artificial sweeteners (NAS) are often chosen to combat weight problems, because they do not contribute to overall calorie intake and are thought to subvert the rise in blood-glucose levels that occurs in response to food intake¹. For unknown reasons, however, NAS are not always effective for weight loss. In this issue, Suez *et al.*² (page 181) describe an unexpected effect of NAS that may shed some light on this issue.

Suez and colleagues added an NAS supplement (saccharin, sucralose or aspartame) to the diets of mice, and found that the sweeteners altered the animals' metabolism, raising blood glucose to significantly higher levels than those of sugar-consuming mice. This was true both for mice fed a normal diet and for those on a high-fat diet — a model for a situation in which NAS supplements might be used to control weight. Because variations in diet have been shown to directly lead to changes in the populations of bacteria that occupy the gut, the authors examined whether these bacteria were responsible for the metabolic changes that they observed. And, indeed, when they used antibiotics to deplete the gut bacteria, they found that this eliminated NAS-induced glucose intolerance in mice fed either diet.

Next, the researchers transplanted faeces from NAS-fed or glucose-fed mice into germ-free mice (those with no gut bacteria of their own) that had never consumed NAS. Transfer from NAS-fed mice induced elevated blood-glucose levels in the transplant recipients. Furthermore, the composition of the recipients'

gut bacterial community was different from that of mice receiving transplants from glucose-fed mice, suggesting that changes in this gut microbiota mediate glucose intolerance in NAS-fed mice. Genetic analysis revealed that this altered composition was accompanied by changes in bacterial function. In particular, Suez and co-workers detected an increase in carbohydrate-degradation pathways in the microbiota of NAS-fed mice. This connection parallels a previous report² that the microbiota of obese mice has a higher carbohydrate-metabolizing capacity than the microbiota of normal-weight mice.

What is the relevance of these results for human disease? Suez *et al.* studied around 400 people, and found that bacterial populations in the guts of those who consumed NAS were significantly different from those who did not. Moreover, NAS consumption correlated with disease markers linked to obesity, such as elevated fasting blood-glucose levels and impaired glucose tolerance.

The authors placed seven volunteers who did not normally consume NAS on a seven-day regimen of controlled high NAS intake. After only four days, half the individuals had elevated blood-glucose levels and altered bacterial-community composition, mirroring the results seen in the mice. Transfer of faeces from NAS-fed

human donors induced elevated blood-glucose levels in germ-free mouse recipients that had never consumed NAS. Taken together, Suez and colleagues' data indicate that NAS consumption may contribute to, rather than alleviate, obesity-related metabolic conditions, by altering the composition and function of bacterial populations in the gut.

Bacterial populations in the guts of people who consumed sweeteners were significantly different from populations of those who did not.

Studies examining genetic^{3,4} and diet-induced⁵ mouse models of obesity and obesity in humans^{6,7} have demonstrated that the disease is associated with changes in the composition of the gut microbiota. Most bacteria colonizing the gut come from two phyla, Bacteroidetes and Firmicutes. Obese mice and humans both have reduced bacterial diversity, with reduced proportions of Bacteroidetes and increased Firmicutes, when compared to lean littermates or twin controls^{4–7}. Obesity-induced changes in the microbiota can be reversed by diet — obese mice or humans on fat- or carbohydrate-restricted diets have an increased abundance of Bacteroidetes^{5,6}.

It is difficult to directly compare Suez and colleagues' findings with earlier work, because the current report describes changes in a mix of bacteria (including Bacteroidetes and Firmicutes) after NAS treatment. Certain gut bacteria are well adapted to break down dietary components that the human body cannot. It could be that expansion of these populations in response to NAS increases extraction of energy — often stored as fat — from the diet, contributing to obesity². Alternatively, NAS might exert their effect by suppressing the growth of particular bacterial taxa. In obese mice, the growth of certain bacterial species is suppressed, and there is an increased production of metabolites that can contribute to insulin resistance⁸. These two possibilities cannot be distinguished in the current report.

Bacterial communities in the gut have been linked to elevated lipid production, and increased storage of lipids and the carbohydrate glycogen^{9,10}, correlating with an increase in adiposity and in cellular energy extraction from food. Furthermore, obesity-induced alterations to the composition of gut microbiota are associated with metabolic changes^{2,8}, including enrichment of pathways related to bacterial growth. This suggests that obesity maintains alterations to the microbiota, allowing for the continued increase in production and storage of lipids and glycogen, further exacerbating the condition. Future work must determine whether the changes in the microbiota brought about by NAS consumption activate any of the same molecular pathways as are active in the obese microbiota.

Type 2 diabetes and impaired glucose tolerance have also been linked to alterations in gut microbiota composition^{11,12}. Analysis of gut bacterial genomes shows that microbial-gene signatures differ between patients with and without diabetes, and people with impaired glucose tolerance^{11,12}. Whether the bacterial populations or metabolic pathways altered

by the consumption of NAS are similar to those described in people with or developing diabetes remains to be seen.

Many diseases associated with Western lifestyles have now been linked to environmentally induced alterations in the composition of the gut microbiota¹³. Questions remain regarding the precise mechanisms by which NAS disrupt the relationship between gut bacteria and their host. Studies to identify specific bacterial populations that promote resistance to weight gain or improve glucose tolerance may prove

useful for devising therapies that modulate bacteria or their metabolites. ■

Taylor Feehley and Cathryn R. Nagler are in the Department of Pathology, The University of Chicago, Chicago, Illinois 60637, USA. e-mail: cnagler@bsd.uchicago.edu

1. Gardner, C. *et al.* *Diabetes Care* **35**, 1798–1808 (2012).
2. Suez, J. *et al.* *Nature* **514**, 181–186 (2014).
3. Turnbaugh, P. J. *et al.* *Nature* **444**, 1027–1031 (2006).
4. Ley, R. E. *et al.* *Proc. Natl Acad. Sci. USA* **102**, 11070–11075 (2005).
5. Turnbaugh, P. J., Bäckhed, F., Fulton, L. & Gordon, J. I.

- Cell Host Microbe* **3**, 213–223 (2008).
- Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. *Nature* **444**, 1022–1023 (2006).
- Turnbaugh, P. J. *et al.* *Nature* **457**, 480–484 (2009).
- Ridaura, V. K. *et al.* *Science* **341**, 1241214 (2013).
- Bäckhed, F. *et al.* *Proc. Natl Acad. Sci. USA* **101**, 15718–15723 (2004).
- Bäckhed, F., Manchester, J. K., Semenkovich, C. F. & Gordon, J. I. *Proc. Natl Acad. Sci. USA* **104**, 979–984 (2007).
- Qin, J. *et al.* *Nature* **490**, 55–60 (2012).
- Karlsson, F. H. *et al.* *Nature* **498**, 99–103 (2013).
- Cho, I. & Blaser, M. J. *Nature Rev. Genet.* **13**, 260–270 (2012).

This article was published online on 17 September 2014.

MATERIALS ANALYSIS

Good vibrations

A newly constructed electron-energy monochromator for an atomic-resolution transmission electron microscope has resolved spectroscopic signatures of chemical-bond vibrations that are spatially highly localized. SEE LETTER P.209

RIK BRYDSON

It is extremely rare for developments in instrumentation to deliver an order of magnitude improvement in resolution or sensitivity for a particular analytical technique, while maintaining existing performance in other areas. A study by Krivanek *et al.*¹ on page 209 of this issue, which focuses on high-energy-resolution electron-energy-loss spectroscopy (EELS) in a scanning transmission electron microscope (STEM), achieves just that. The results promise a new territory for experimental investigation of both fundamental physics and the structure of advanced materials at the atomic scale.

The work is a true collaboration between a small manufacturer of electron microscopes and several academics, including teams from two US universities that have recently installed this STEM/EELS instrumentation. It represents a good example of a research community working together to drive forward a particular technique. The paper describes the practical implementation of a design^{2,3} for an electron monochromator — a device that produces an electron beam that has a narrow electron-energy distribution. The electron monochromator acts on the high-energy electron beam of a STEM, which is corrected for aberrations in the lens that focuses the electron beam into a small probe of atomic dimensions⁴.

In a STEM, a high-energy electron probe is transmitted through a thin specimen, undergoing energy losses as a result of scattering from atoms in the sample. The probe is scanned over the specimen to form images of it as a function of probe position. If an electron spectrometer is added to a STEM, a spectrum of the transmitted electron intensity versus energy loss — the EELS data — is

also obtained. Krivanek and colleagues show that, with the inclusion of appropriate energy-stabilization schemes, the monochromator design that they have implemented can achieve an electron-energy resolution for EELS in a STEM of around 10 millielectronvolts (Fig. 1), while retaining a sufficiently small probe diameter with enough electron current to maintain atomic-resolution imaging and analysis.

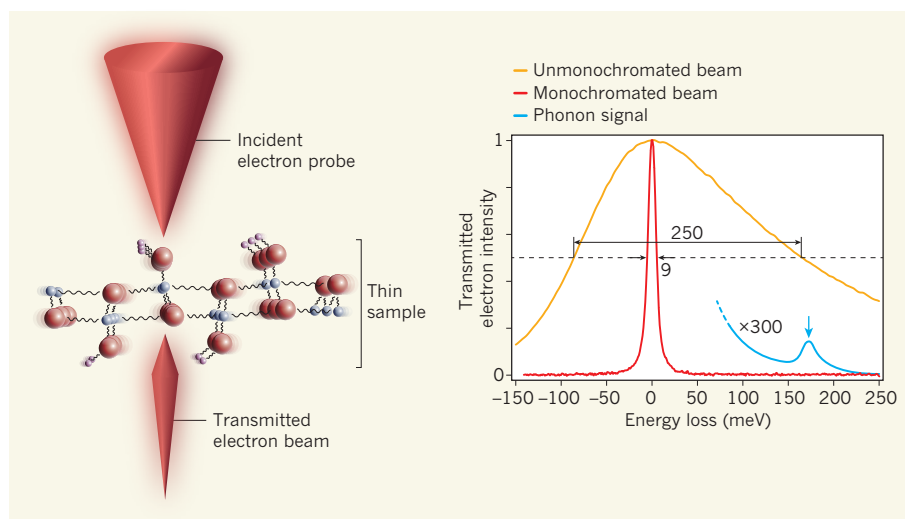


Figure 1 | Spatially resolved vibrational spectroscopy. In a scanning transmission electron microscope corrected for aberration effects, a beam of high-energy electrons is focused onto a thin sample, forming a probe of atomic dimensions. The electron beam transmitted through the sample, which in this case comprises different types of atom and bond length, contains electrons that have experienced energy losses owing to scattering events with atoms in the sample. These electrons can be dispersed in terms of their energy using a magnetic prism spectrometer (not shown), enabling a spectrum of scattered electron intensity versus energy loss to be obtained. Krivanek *et al.*¹ show that, if the incident electron beam goes through an energy monochromator (not shown), then the energy spread associated with a beam that undergoes zero energy loss can be improved from 250 millielectronvolts to 9 meV. This improvement allows the detection of spectroscopic signatures of spatially localized chemical-bond vibrations (phonons). The intensity scale of the phonon signal is magnified by a factor of 300 to show the phonon peak (blue arrow). (Plot adapted from ref. 1.)

These are out-of-phase lattice vibrations in the plane of the thin sample, arising when neighbouring atoms in the lattice have different charge or mass.

Until now, lattice vibrations were something electron microscopists have had to worry about only in terms of the sample damage that they induce, or when matching experimental images to simulations⁷. However, the main point of Krivanek and colleagues' work is that optical phonons are key signatures of chemical bonds, particularly those involving light elements such as hydrogen, as is well established by the techniques of infrared and Raman (optical) spectroscopy. The implication is, therefore, that STEM-EELS may provide a route for the direct mapping of chemical bonding, including that associated with light elements, at near-atomic resolution.

This achievement would present tremendous benefits in a number of highly topical areas of research into new advanced materials and devices. The improvements in overall energy resolution of EELS will undoubtedly aid the study of the local spatial variation of energy bandgaps in semiconducting structures, and the identification of localized collective oscillation of electrons in 'plasmonic' structures for light capture. The ability to detect and map light elements, including hydrogen, could extend the existing capability of analytical electron microscopy to the study of organic materials such as polymers and pharmaceuticals, as well as energy-storage materials — if issues associated with electron-beam-induced damage can be addressed. Directly measuring phonons could potentially help to identify chemical reactions involving the surfaces of nanoscale heterogeneous catalyst particles, and could aid the investigation of the transmission of lattice vibrations across micro- and nanostructural features, such as interfaces and defects in thermal and optical materials. As with the emergence of any new technique, many additional research areas may ultimately prove to be most fruitful.

Existing theory suggests that electrons that have undergone phonon scattering would be scattered through large angles and the resulting phonon signal would be spatially highly delocalized, preventing atomic-resolution analysis. However, Krivanek and colleagues present some initial findings which, together with recent theoretical predictions⁸, suggest that under appropriate conditions the phonon signal may be sufficiently localized for the study of vibrations at a spatial resolution better than that achieved by scanning probe tip-enhanced vibrational spectroscopies⁹. The authors observed an exponential delocalization of the phonon signal as an electron probe is moved away from the surface of a sample and into the surrounding vacuum. However, there seems to be a more localized component of the signal that peaks in intensity close to the surface itself, and the researchers discuss a possible experimental geometry for signal collection that would

enhance this more localized contribution. Furthermore, if the probe is inside the sample, it seems that the delocalization could be screened at the interface between two materials with different electrical properties.

The authors also demonstrate a method for remotely exciting such phonons at a surface using the inherent delocalization of the signal; here, the beam is located in the vacuum close to the edge of a sample, potentially helping to mitigate electron-beam-induced damage of radiation-sensitive samples. These investigations of the spatial resolution of the phonon signal represent a clear example of experiment leading theory in terms of the interpretation of the results, and is indicative of the new experimental landscape that this development in instrumentation unfolds. Undoubtedly, many more exciting experiments with this technology will follow, aided by the delivery, later in 2014, of a third-generation instrument to a shared user facility: the Engineering and Physical Sciences Research Council National

Facility for Aberration-Corrected STEM, or SuperSTEM¹⁰, in Daresbury, UK. I look forward to the community charting this new frontier of research. ■

Rik Brydson is at the Institute for Materials Research, School of Chemical and Process Engineering, University of Leeds, Leeds LS2 9JT, UK.

e-mail: r.m.drummond-brydson@leeds.ac.uk

1. Krivanek, O. L. *et al.* *Nature* **514**, 209–212 (2014).
2. Krivanek, O. L. *et al.* *Phil. Trans. R. Soc. A* **367**, 3683–3697 (2009).
3. Krivanek, O. L. *et al.* *Microscopy* **62**, 3–21 (2013).
4. Bleloch, A. & Ramasse, Q. in *Aberration-Corrected Analytical Transmission Electron Microscopy* (ed. Brydson, R.) Ch. 4, 55–87 (Wiley, 2011).
5. Egerton, R. F. *Electron Energy Loss Spectroscopy in the Electron Microscope* 3rd edn (Springer, 2011).
6. Mahan, G. D. *Condensed Matter in a Nutshell* (Princeton Univ. Press, 2010).
7. Williams, D. B. & Carter, C. B. *Transmission Electron Microscopy* 2nd edn (Springer, 2009).
8. Dwyer, C. *Phys. Rev. B* **89**, 054103 (2014).
9. Hermann, P. *et al.* *Analyst* **136**, 1148–1152 (2011).
10. www.superstem.org

ANIMAL BEHAVIOUR

Incipient tradition in wild chimpanzees

The adoption of a new form of tool use has been observed to spread along social-network pathways in a chimpanzee community. The finding offers the first direct evidence of cultural diffusion in these animals in the wild.

ANDREW WHITEN

Social learning — learning from others — is one of the fastest-expanding research fields in animal behaviour^{1,2}. At the fundamental level of evolutionary biology, social learning provides a high-speed 'second inheritance system' that interacts with genetic inheritance to enrich behavioural evolution². From a more anthropocentric perspective, animal social learning casts light on the evolutionary foundations of the cultural capacities that make our own species so successful. Studies of putative cultural variations in wild chimpanzees^{2–4}, the primates with which (together with bonobos) humans last shared a common ancestor, have been particularly influential in our understanding of behavioural evolution. But these observed regional behavioural differences have displayed little change, making it difficult to investigate the workings of social learning. Now, writing in *PLoS Biology*, Hobaiter *et al.*⁵ describe a new form of tool use in the Sonso community of chimpanzees of the Budongo Forest in Uganda, and present a statistical technique for tracing the social transmission of this innovation.

Chimpanzees at Sonso fold wads of leaves in their mouths to fashion a 'leaf sponge' that they dip into tree holes to extract water to drink. In 2011, the researchers observed the dominant male of the group, Nick, creating a sponge of moss gathered from a tree trunk and using it to drink from a small flooded waterhole — a behaviour not previously recorded in the 20-year research programme at the site. He was watched by the dominant female, Nambi. Over the next six days of intensive and often competitive use of the waterhole (which the researchers suspect contained unusual densities of minerals or other desirable content), Nambi and six other chimpanzees began to display the moss-sponging technique (Fig. 1a). More than 20 other individuals drank at the hole or in puddles around it, but either directly with their mouths or using leaf sponges rather than moss sponges.

To establish whether this behavioural spread was due to social learning, the researchers developed a form of network-based diffusion analysis (NBDA). This statistical technique quantifies the extent to which the spread of a new behaviour is consistent with the prediction that it will follow the social network — a

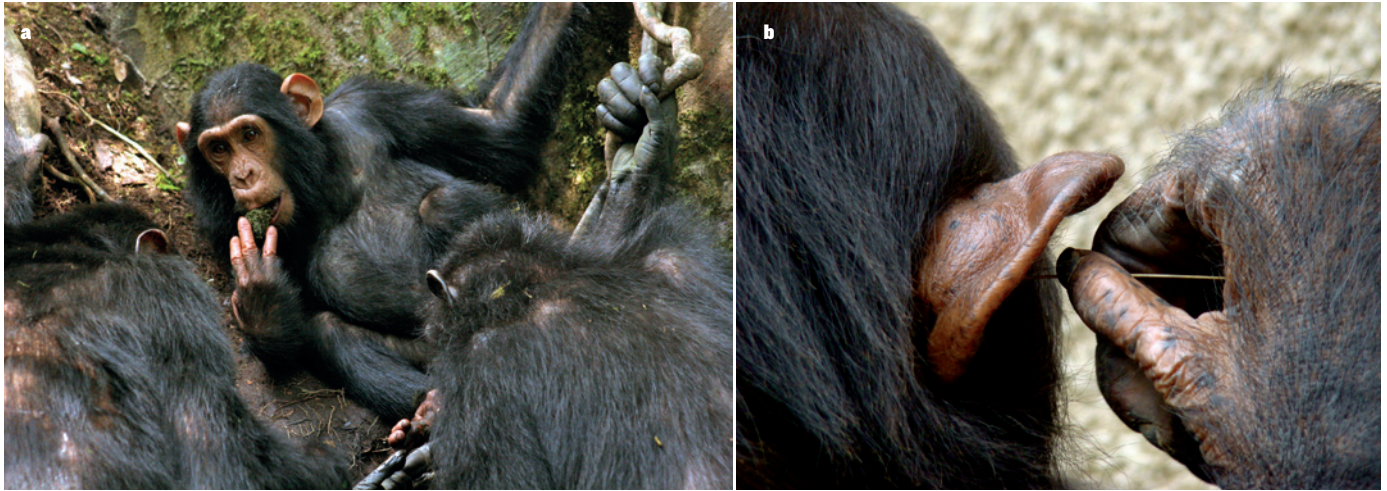


Figure 1 | Tools and trends. **a**, A chimpanzee from the Sonso community using a moss sponge to drink from a waterhole. Hobaite *et al.*⁵ observed that this new tool use spread from one animal to others along the community's social network. **b**, Another study⁹ reports the spread of a new and seemingly useless behavioural 'fad' — sticking a blade of grass in the ear — among chimpanzees in a captive sanctuary community.

numerical representation of who associates most closely with whom in the community. An impressive recent example of NBDA⁶ used more than 73,000 observations of 653 humpback whales during the 27-year spread of a 'lobtail' prey-capture technique, which diffused as predicted by the social network, implying social learning. However, NBDA studies have so far used only a static, summary quantification of the social network. Hobaite *et al.* took this approach to a new level, which they describe as dynamic NBDA, by incorporating repeatedly updated information on whom each individual was likely to have watched (those within 1 metre of and facing the current sponger). They found strong evidence consistent with social transmission, with an estimated 15-fold enhancement of moss-sponging behaviour for each time a novice observed an existing moss-sponger.

Of course, NBDA and other purely statistical approaches to analysing observational data are essentially correlational and thus do not necessarily imply cause. To interpret such data, one has to try to rule out alternative potential explanations — in this case, for example, that the order of acquisition observed in the chimpanzees resulted not from social learning but from rank-based queuing to gain access to the waterhole, with each lower-ranked individual happening to watch the previous higher ranker before they got their turn. It seems that this possibility can be excluded, because the 'lower rankers' were often offspring of the higher-ranked earlier moss-spongers and so had simultaneous access to the waterhole, yet started to use moss only after watching their mothers. However, caution is still warranted in interpreting these findings in case some subtle alternative factor explains the observed putative evidence for social learning.

Experimental approaches can provide more-robust tests of causality. Indisputable cases of social learning have already been seen

in captive primates, including chimpanzees, in studies in which alternative techniques for using tools or otherwise manipulating foraging tasks are seeded and subsequently spread in different groups⁷. Such approaches are inherently difficult to engineer in the field, but a few attempts have been made (see ref. 8 for a review). Unfortunately, this method has yet to be successfully implemented with wild chimpanzees, which are surprisingly neophobic.

Nevertheless, non-interventional studies of natural behaviour, such as the one presented by Hobaite *et al.*, are vital to the field. Experimental studies make good sense only when they build on what has first been established in the wild. The innovation recorded by the authors was not dramatic — it was merely a modification of existing leaf-sponging expertise. But the findings are valuable as the first direct evidence of cultural diffusion in this key species, converging with observational evidence from the wild and rigorously controlled experiments in captive animals to consolidate a substantial case for the role of cultural transmission in such cases.

This study follows hot on the heels of another⁹ documenting the diffusion of a particularly intriguing innovation — placing a blade of grass in the ear — in chimpanzees living in four large enclosures in an African sanctuary (Fig. 1b). That study is unusual because the behaviour seems to be functionless, and thus akin to human cultural phenomena such as fads and fashions. The grass-in-ear behaviour spread from one apparent inventor in 2007 to seven others by 2012, in just one of the four groups. The lack of overt function makes any explanation other than social learning difficult to accept, and underlines the potential potency of this form of learning in this species.

Researchers have also claimed the first documented case of successful transmission of a novel cultural behaviour — fishing for wood-boring ants using peeled bark or

other material — between wild chimpanzee communities¹⁰. And in other studies of the selection of materials for nut-cracking, researchers concluded that migrating female chimpanzees soon conform to the practices of the group they move into¹¹. A major question for the future is thus what determines the outcome of such migrations between different local cultures. Why do migrants sometimes seed behaviours that diffuse in their new community in the manner demonstrated in the moss-sponging study, whereas others instead abandon previous behaviours and conform to the new local norms? Investigating which factors throw this important switch will add considerably to our understanding of cultural transmission in animals. ■

Andrew Whiten is at the Centre for Social Learning and Cognitive Evolution, School of Psychology and Neuroscience, University of St Andrews, St Andrews KY16 9JP, UK.
e-mail: a.whiten@st-andrews.ac.uk

- Hoppitt, W. & Laland, K. N. *Social Learning: An Introduction to Mechanisms, Methods, and Models* (Princeton Univ. Press, 2013).
- Whiten, A. *Nature* **437**, 52–55 (2005).
- Nishida, T., Zamma, K., Matsusaka, T., Inaba, A. & McGrew, W. C. *Chimpanzee Behavior in the Wild: An Audio-Visual Encyclopedia* (Springer, 2010).
- Boesch, C. *Wild Cultures: A Comparison between Chimpanzee and Human Cultures* (Cambridge Univ. Press, 2012).
- Hobaite, C., Poisot, T., Zuberbühler, K., Hoppitt, W. & Gruber, T. *PLoS Biol.* **12**, e1001960 (2014).
- Allen, J., Weinrich, M., Hoppitt, W. & Rendell, L. *Science* **340**, 485–488 (2013).
- Whiten, A. & Mesoudi, A. *Phil. Trans. R. Soc. B* **363**, 3477–3488 (2008).
- van de Waal, E., Borgeaud, C. & Whiten, A. *Science* **340**, 483–485 (2013).
- van Leeuwen, E. J. C., Cronin, K. A. & Haun, D. B. M. *Anim. Cogn.* <http://dx.doi.org/10.1007/s10071-014-0766-8> (2014).
- O'Malley, R. C., Wallauer, W., Murray, C. M. & Goodall, J. *Curr. Anthropol.* **53**, 650–663 (2012).
- Luncz, L. V. & Boesch, C. *Am. J. Primatol.* **76**, 649–657 (2014).

This article was published online on 1 October 2014.

Artificial sweeteners induce glucose intolerance by altering the gut microbiota

Jotham Suez¹, Tal Korem^{2*}, David Zeevi^{2*}, Gili Zilberman-Schapira^{1*}, Christoph A. Thaiss¹, Ori Maza¹, David Israeli³, Niv Zmora^{4,5,6}, Shlomit Gilad⁷, Adina Weinberger², Yael Kuperman⁸, Alon Harmelin⁸, Ilana Kolodkin-Gal⁹, Hagit Shapira¹, Zamir Halpern^{5,6}, Eran Segal² & Eran Elinav¹

Non-caloric artificial sweeteners (NAS) are among the most widely used food additives worldwide, regularly consumed by lean and obese individuals alike. NAS consumption is considered safe and beneficial owing to their low caloric content, yet supporting scientific data remain sparse and controversial. Here we demonstrate that consumption of commonly used NAS formulations drives the development of glucose intolerance through induction of compositional and functional alterations to the intestinal microbiota. These NAS-mediated deleterious metabolic effects are abrogated by antibiotic treatment, and are fully transferrable to germ-free mice upon faecal transplantation of microbiota configurations from NAS-consuming mice, or of microbiota anaerobically incubated in the presence of NAS. We identify NAS-altered microbial metabolic pathways that are linked to host susceptibility to metabolic disease, and demonstrate similar NAS-induced dysbiosis and glucose intolerance in healthy human subjects. Collectively, our results link NAS consumption, dysbiosis and metabolic abnormalities, thereby calling for a reassessment of massive NAS usage.

Non-caloric artificial sweeteners (NAS) were introduced over a century ago as means for providing sweet taste to foods without the associated high energy content of caloric sugars. NAS consumption gained much popularity owing to their reduced costs, low caloric intake and perceived health benefits for weight reduction and normalization of blood sugar levels¹. For these reasons, NAS are increasingly introduced into commonly consumed foods such as diet sodas, cereals and sugar-free desserts, and are being recommended for weight loss and for individuals suffering from glucose intolerance and type 2 diabetes mellitus¹.

Some studies showed benefits for NAS consumption² and little induction of a glycaemic response³, whereas others demonstrated associations between NAS consumption and weight gain⁴, and increased type 2 diabetes risk⁵. However, interpretation is complicated by the fact that NAS are typically consumed by individuals already suffering from metabolic syndrome manifestations. Despite these controversial data, the US Food and Drug Administration (FDA) approved six NAS products for use in the United States.

Most NAS pass through the human gastrointestinal tract without being digested by the host^{6,7} and thus directly encounter the intestinal microbiota, which plays central roles in regulating multiple physiological processes⁸. Microbiota composition⁹ and function¹⁰ are modulated by diet in the healthy/lean state as well as in obesity^{11,12} and diabetes mellitus¹³, and in turn microbiota alterations have been associated with propensity to metabolic syndrome¹⁴. Here, we study NAS-mediated modulation of microbiota composition and function, and the resultant effects on host glucose metabolism.

Chronic NAS consumption exacerbates glucose intolerance

To determine the effects of NAS on glucose homeostasis, we added commercial formulations of saccharin, sucralose or aspartame to the

drinking water of lean 10-week-old C57Bl/6 mice (Extended Data Fig. 1a). Since all three commercial NAS comprise ~5% sweetener and ~95% glucose, we used as controls mice drinking only water or water supplemented with either glucose or sucrose. Notably, at week 11, the three mouse groups that consumed water, glucose and sucrose featured comparable glucose tolerance curves, whereas all three NAS-consuming mouse groups developed marked glucose intolerance ($P < 0.001$, Fig. 1a, b).

As saccharin exerted the most pronounced effect, we further studied its role as a prototypical artificial sweetener. To corroborate the findings in the obesity setup, we fed C57Bl/6 mice a high-fat diet (HFD, 60% kcal from fat) while consuming either commercial saccharin or pure glucose as a control (Extended Data Fig. 1b). As in the lean state, mice fed HFD and commercial saccharin developed glucose intolerance, compared to the control mouse group ($P < 0.03$, Fig. 1c and Extended Data Fig. 2a). To examine the effects of pure saccharin on glucose intolerance, we followed a cohort of 10-week-old C57Bl/6 mice fed on HFD and supplemented with 0.1 mg ml⁻¹ of pure saccharin added to their drinking water (Extended Data Fig. 1c). This dose corresponds to the FDA acceptable daily intake (ADI) in humans (5 mg per kg (body weight), adjusted to mouse weights, see Methods). As with commercial saccharin, this lower dose of pure saccharin was associated with impaired glucose tolerance ($P < 0.0002$, Fig. 1d and Extended Data Fig. 2b) starting as early as 5 weeks after HFD initiation. Similarly, HFD-fed outbred Swiss Webster mice supplemented with or without 0.1 mg ml⁻¹ of pure saccharin (Extended Data Fig. 1d) showed significant glucose intolerance after 5 weeks of saccharin exposure as compared to controls ($P < 0.03$, Extended Data Fig. 2c, d).

Metabolic profiling of normal-chow- or HFD-fed mice in metabolic cages, including liquids and chow consumption, oxygen consumption, walking distance and energy expenditure, showed similar measures between NAS- and control-drinking mice (Extended Data Fig. 3 and 4).

¹Department of Immunology, Weizmann Institute of Science, Rehovot 76100, Israel. ²Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel. ³Day Care Unit and the Laboratory of Imaging and Brain Stimulation, Kfar Shaul hospital, Jerusalem Center for Mental Health, Jerusalem 91060, Israel. ⁴Internal Medicine Department, Tel Aviv Sourasky Medical Center, Tel Aviv 64239, Israel. ⁵Research Center for Digestive Tract and Liver Diseases, Tel Aviv Sourasky Medical Center, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 69978, Israel. ⁶Digestive Center, Tel Aviv Sourasky Medical Center, Tel Aviv 64239, Israel. ⁷The Nancy and Stephen Grand Israel National Center for Personalized Medicine (INCPM), Weizmann Institute of Science, Rehovot 76100, Israel.

⁸Department of Veterinary Resources, Weizmann Institute of Science, Rehovot 76100, Israel. ⁹Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel.

*These authors contributed equally to this work.

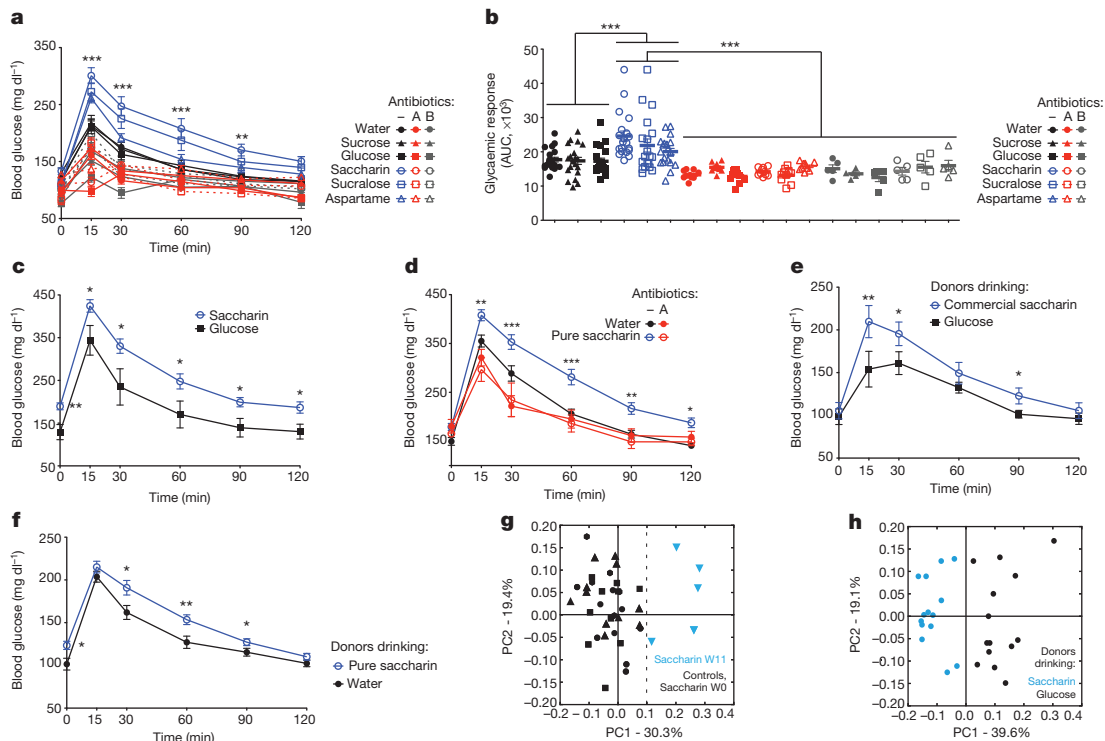


Figure 1 | Artificial sweeteners induce glucose intolerance transferable to germ-free mice. **a, b**, Oral glucose tolerance test (OGTT, **a**) and area under the two-hour blood glucose response curve (AUC, **b**) in normal-chow-fed mice drinking commercial NAS for 11 weeks before ($N = 20$) and after antibiotics: ciprofloxacin and metronidazole ('antibiotics A', $N = 10$); or vancomycin ('antibiotics B', $N = 5$). **c**, OGTT in mice fed HFD and commercial saccharin ($N = 10$) or glucose ($N = 9$). **d**, OGTT of HFD-fed mice drinking 0.1 mg ml⁻¹ saccharin or water for 5 weeks ($N = 20$), followed by 'antibiotics A' ($N = 10$). **e, f**, OGTT of germ-free mice 6 days following transplant of microbiota from commercial saccharin- ($N = 12$) and glucose-fed mice ($N = 11$) (**e**), or pure

saccharin- ($N = 16$) and water-fed ($N = 16$) donors (**f**). Symbols (OGTT) or horizontal lines (AUC), mean; error bars, s.e.m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. OGTT, analysis of variance (ANOVA) and Bonferroni; AUC, ANOVA and Tukey post hoc analysis. Each experiment was repeated twice. **g**, Principal coordinates analysis (PCoA) of weighted UniFrac distances based on 16S rRNA analysis from saccharin-consuming mice at baseline (W0, black hexagons; W11, blue triangles); water controls (black circles for W11 and W0); glucose (black squares for W11 and W0); or sucrose (black triangles for W11 and W0). $N = 5$ in each group. **h**, PCoA of taxa in germ-free recipients according to donor identity in **e**.

Fasting serum insulin levels and insulin tolerance were also similar in all mouse groups consuming NAS or caloric sweeteners, in both the normal-chow and HFD settings (Extended Data Fig. 5). Taken together, these results suggest that NAS promote metabolic derangements in a range of formulations, doses, mouse strains and diets paralleling human conditions, in both the lean and the obese state.

Gut microbiota mediates NAS-induced glucose intolerance

Since diet modulates the gut microbiota¹⁵, and microbiota alterations exert profound effects on host physiology and metabolism, we tested whether the microbiota may regulate the observed NAS effects. To this end, we treated mouse groups consuming commercial or pure NAS in the lean and HFD states (Extended Data Fig. 1a, c) with a Gram-negative-targeting broad-spectrum antibiotics regimen (designated 'antibiotics A') of ciprofloxacin (0.2 g l⁻¹) and metronidazole (1 g l⁻¹), while maintaining mice on their diet and sweetener supplementation regimens. Notably, after 4 weeks of antibiotic treatment, differences in glucose intolerance between NAS-drinking mice and controls were abolished both in the lean (Fig. 1a, b) and the obese (Fig. 1d and Extended Data Fig. 2b) states. Similar effects were observed with the Gram-positive-targeting antibiotic vancomycin ('antibiotics B', 0.5 g l⁻¹, Fig. 1a, b). These results suggest that NAS-induced glucose intolerance is mediated through alterations to the commensal microbiota, with contributions from diverse bacterial taxa.

To test whether the microbiota role is causal, we performed faecal transplantation experiments, by transferring the microbiota configuration from mice on normal-chow diet drinking commercial saccharin

or glucose (control) into normal-chow-consuming germ-free mice (Extended Data Fig. 1e). Notably, recipients of microbiota from mice consuming commercial saccharin exhibited impaired glucose tolerance as compared to control (glucose) microbiota recipients, determined 6 days following transfer ($P < 0.03$, Fig. 1e and Extended Data Fig. 2e). Transferring the microbiota composition of HFD-consuming mice drinking water or pure saccharin replicated the glucose intolerance phenotype ($P < 0.004$, Fig. 1f and Extended Data Fig. 2f). Together, these results establish that the metabolic derangements induced by NAS consumption are mediated by the intestinal microbiota.

NAS mediate distinct functional alterations to the microbiota

We next examined the faecal microbiota composition of our various mouse groups by sequencing their 16S ribosomal RNA gene. Mice drinking saccharin had a distinct microbiota composition that clustered separately from both their starting microbiome configurations and from all control groups at week 11 (Fig. 1g). Likewise, microbiota in germ-free recipients of stools from saccharin-consuming donor mice clustered separately from that of germ-free recipients of glucose-drinking donor stools (Fig. 1h). Compared to all control groups, the microbiota of saccharin-consuming mice displayed considerable dysbiosis, with more than 40 operational taxonomic units (OTUs) significantly altered in abundance (false discovery rate (FDR) corrected P value < 0.05 for each OTU; Extended Data Fig. 6, Supplementary Table 1). Many of the taxa that increased in relative abundance belonged to the *Bacteroides* genus and Clostridiales order, with other members of the Clostridiales order comprising the majority of under-represented taxa, along with *Lactobacillus reuteri*, and were

mirrored in germ-free recipients of stools from saccharin-consuming donors (Extended Data Fig. 6, right column). Likewise, dysbiosis was observed in mice consuming pure saccharin and HFD (Supplementary Table 1). Together, these results demonstrate that saccharin consumption in various formulations, doses and diets induces dysbiosis with overall similar configurations.

To study the functional consequences of NAS consumption, we performed shotgun metagenomic sequencing of faecal samples from before and after 11 weeks of commercial saccharin consumption, compared to control mice consuming either glucose or water. To compare relative species abundance, we mapped sequencing reads to the human microbiome project reference genome database¹⁶. In agreement with the 16S rRNA analysis, saccharin treatment induced the largest changes in microbial relative species abundance (Fig. 2a, Supplementary Table 2; F-test P value $< 10^{-10}$). These changes are unlikely to be an artefact of horizontal gene transfer or poorly covered genomes, because changes in relative abundance were observed across much of the length of the bacterial genomes, as exemplified by one overrepresented (*Bacteroides vulgatus*, Extended Data Fig. 7a) and one underrepresented species (*Akkermansia muciniphila*, Extended Data Fig. 7b).

We next mapped the metagenomic reads to a gut microbial gene catalogue, evenly dividing reads mapping to more than one gene, and then grouping genes into KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways. Examining pathways with gene coverage above 0.2 (115 pathways), changes in pathway abundance were inversely correlated between commercial saccharin- and glucose-consuming mice ($R = -0.45$, $P < 10^{-6}$, Fig. 2b). Since commercial saccharin consists of 95% glucose, these results suggest that saccharin greatly affects microbiota function. Notably, pathways overrepresented in saccharin-consuming mice include a strong increase in glycan degradation pathways (Fig. 2c, d), in which glycans are fermented to form various compounds including short-chain fatty acids (SCFAs)¹⁷. These pathways mark enhanced energy harvest and their enrichment was previously associated with obesity in mice¹¹ and humans¹⁸, with SCFA possibly serving as precursors and/or signalling molecules for *de novo* glucose and lipid synthesis by the host¹⁹. To identify the underlying bacteria, we annotated every read that mapped to glycan degradation pathways by its originating bacteria. Much of the increase in these pathways is attributable to reads originating from five Gram-negative and -positive species, of which two belong to the

Bacteroides genus (Fig. 2e). This is consistent with the sharp increase in the abundance of this genus in saccharin-consuming mice observed in the 16S rRNA analysis (Extended Data Fig. 6). Consequently, levels of the SCFAs propionate and acetate measured in stool were markedly higher in commercial saccharin-consuming mice compared to control glucose-consuming mice (Fig. 2f, g), reflective of the differential effects mediated by chronic glucose consumption with and without NAS exposure. Butyrate levels were similar between the groups (data not shown).

In addition to glycan degradation, and similar to previous studies on humans with type 2 diabetes^{13,20}, other pathways were enriched in microbiomes of saccharin-consuming mice, including starch and sucrose metabolism, fructose and mannose metabolism, and folate, glycerolipid and fatty acid biosynthesis (Supplementary Tables 3 and 4), whereas glucose transport pathways were underrepresented in saccharin-consuming mice (Extended Data Fig. 7c). Mice consuming HFD and pure saccharin featured several enriched pathways (Extended Data Fig. 7d), including ascorbate and aldarate metabolism (previously reported to be enriched in leptin-receptor-deficient diabetic mice²¹), lipopolysaccharide biosynthesis (linked to metabolic endotoxemia²²) and bacterial chemotaxis (previously reported to be enriched in obese mice¹¹).

Altogether, saccharin consumption results in distinct diet-dependent functional alterations in the microbiota, including normal-chow-related expansion in glycan degradation contributed by several of the increased taxa, ultimately resulting in elevated stool SCFA levels, characteristic of increased microbial energy harvest¹¹.

NAS directly modulate the microbiota to induce glucose intolerance

To determine whether saccharin directly affects the gut microbiota, we cultured faecal matter from naive mice under strict anaerobic conditions (75% N₂, 20% CO₂, 5% H₂) in the presence of saccharin (5 mg ml⁻¹) or control growth media. Cultures from day 9 of incubation were administered by gavage to germ-free mice (Extended Data Fig. 8a). *In vitro* stool culture with saccharin induced an increase of the Bacteroidetes phylum and reduction in Firmicutes (Bacteroidetes 89% versus 70%, Firmicutes 6% versus 22%, Extended Data Fig. 8b). Transferring this *in vitro* saccharin-treated microbiota configuration into germ-free mice resulted in significantly higher glucose intolerance ($P < 0.002$) compared with germ-free mice receiving the control culture (Fig. 3a and Extended

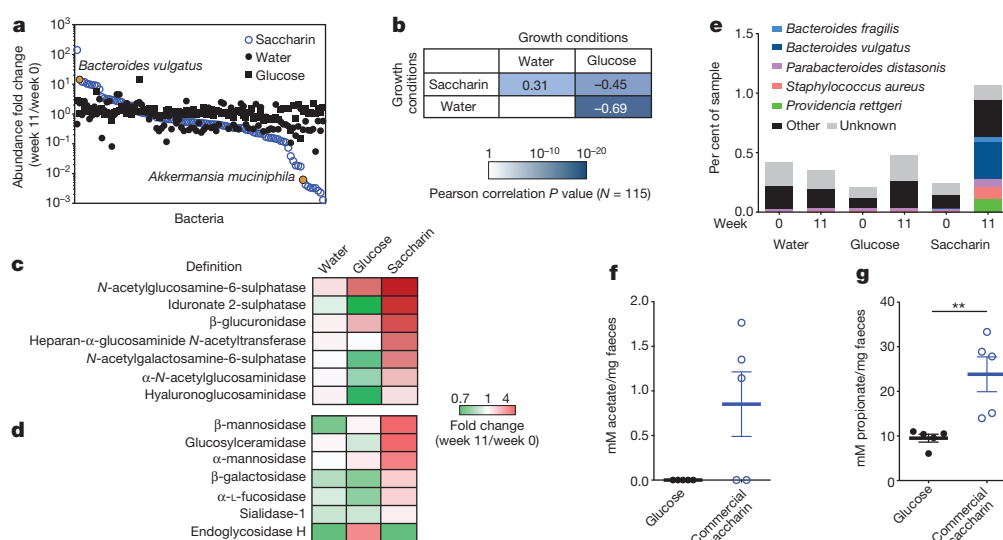


Figure 2 | Functional characterization of saccharin-modulated microbiota. **a**, Species alterations in mice consuming commercial saccharin, water or glucose for 11 weeks ($N = 4$) shown by shotgun sequencing. **b**, Pairwise correlations between changes in 115 KEGG pathways across mice receiving different treatments. **c**, **d**, Fold change in relative abundance of glycosaminoglycan (**c**) or other glycan (**d**) degradation pathway genes.

e, Higher glycan degradation attributed to five taxa in the commercial saccharin setting. **f**, **g**, Levels of faecal acetate and propionate at W11 in mice drinking commercial saccharin or glucose ($N = 5$). Horizontal lines, means; error bars, s.e.m. ** $P < 0.01$, two-sided unpaired Student *t*-test. SCFA measurements were performed on two independent cohorts.

Data Fig. 8c). Similar to the composition of the saccharin-supplemented anaerobic culture, germ-free recipients of this cultured configuration featured over-representation of members of the *Bacteroides* genus, and under-representation of several Clostridiales (Fig. 3b and Supplementary Table 5).

Shotgun metagenomic sequencing analysis revealed that *in vitro* saccharin treatment induced similar functional alterations to those found in mice consuming commercial saccharin (Fig. 4c, $P < 10^{-4}$), with glycan degradation pathways being highly enriched in both settings. Other pathways highly enriched in both settings included those involved in sphingolipid metabolism, previously shown to be over-represented in microbiomes of non-obese diabetic mice²³, and common under-represented pathways included glucose transport (Fig. 3c and Extended Data 7c, right column).

Collectively, these results demonstrate that saccharin directly modulates the composition and function of the microbiome and induces

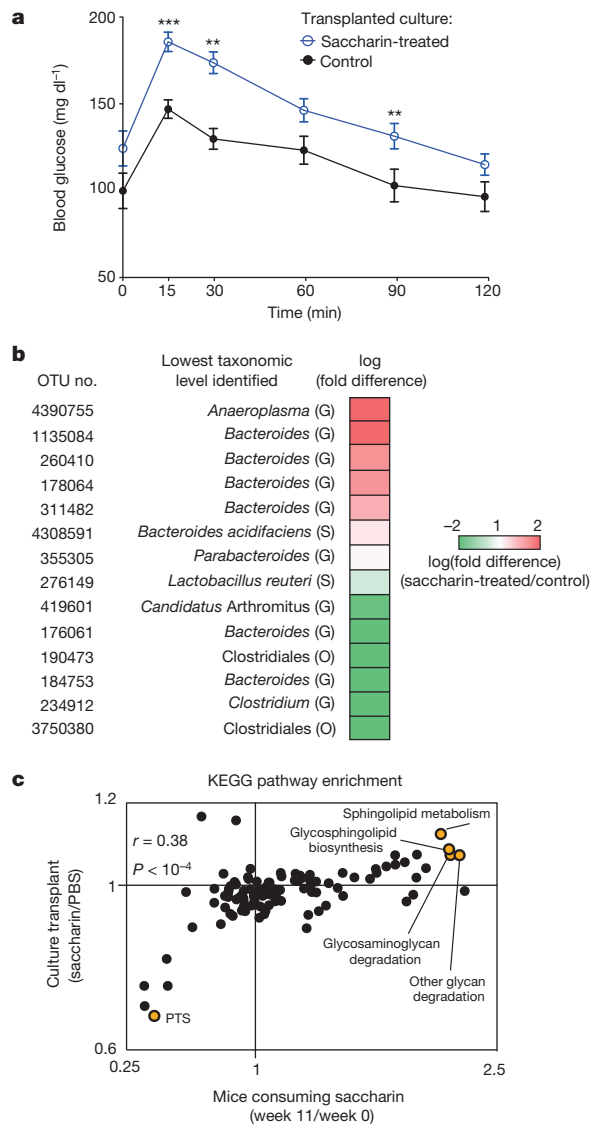


Figure 3 | Saccharin directly modulates the microbiota. **a**, OGTT of germ-free mice 6 days following transplantation with saccharin-enriched or control faecal cultures ($N = 10$ and $N = 9$, respectively). Symbols, means; error bars, s.e.m. *** $P < 0.01$, ** $P < 0.001$, ANOVA and Bonferroni post hoc analysis. Experiments were repeated twice. **b**, Taxa representation in germ-free mice from **a**. O, order; G, genus; S, species. **c**, Comparison of KEGG pathway abundance between W11 saccharin-consuming mice (compared to W0, x axis) and germ-free mice transplanted with microbiomes anaerobically cultured with 5 mg ml⁻¹ saccharin (compared to PBS, y axis).

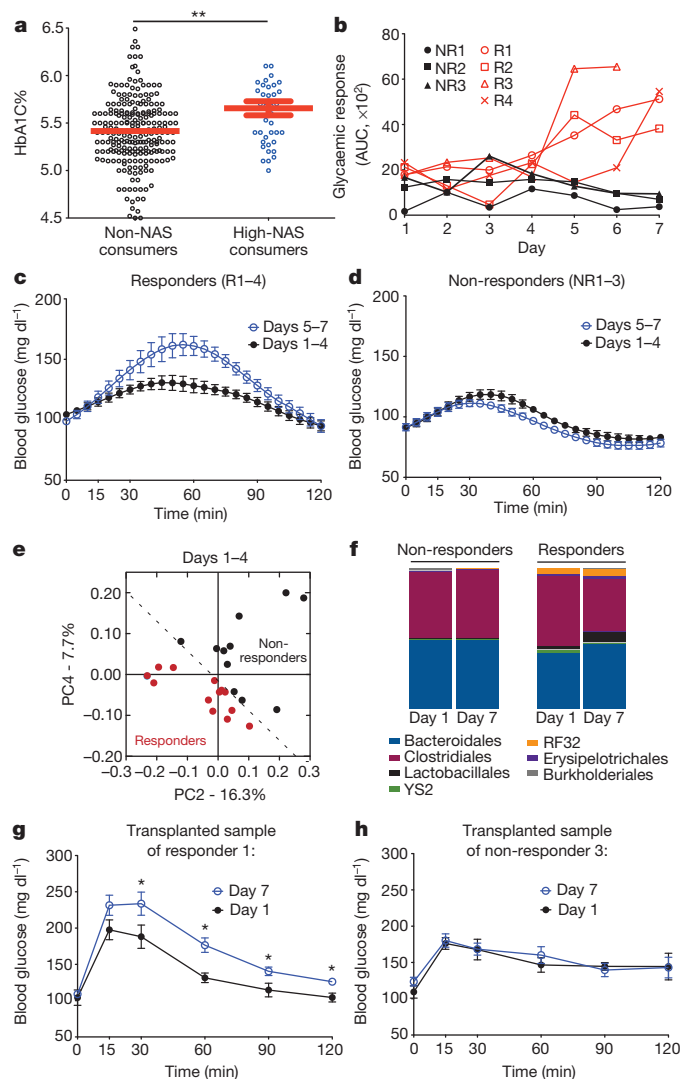


Figure 4 | Acute saccharin consumption impairs glycaemic control in humans by inducing dysbiosis. **a**, HbA1C% of high NAS consumers ($N = 40$) versus non-consumers ($N = 236$). **Rank sum P value < 0.002 . **b**, Daily incremental area under the curve (iAUC) of OGTT in 4 responders (R) and 3 non-responders (NR). **c**, **d**, OGTT of days 1–4 versus days 5–7 in 4 responders (c) or 3 non-responders (d). **e**, PCoA of 16S rRNA sequences from responders (N samples = 16) versus non-responders ($N = 9$) during days 1–4. **f**, Order-level relative abundance of taxa samples from days 1 and 7 of responders and non-responders. **g**, **h**, OGTT in germ-free mice ($N = 6$) 6 days following faecal transplantation of D1 or D7 samples of responder 1 (R1, **g**) or non-responder 3 (NR3, **h**). Symbols, means; error bars, s.e.m. * $P < 0.05$, ** $P < 0.01$, ANOVA and Bonferroni post hoc analysis.

dysbiosis, accounting for the downstream glucose intolerance phenotype in the mammalian host.

NAS in humans associate with impaired glucose tolerance

To study the effect of NAS in humans, we examined the relationship between long-term NAS consumption (based on a validated food frequency questionnaire, see Methods) and various clinical parameters in data collected from 381 non-diabetic individuals (44% males and 56% females; age 43.3 ± 13.2) in an ongoing clinical nutritional study. We found significant positive correlations between NAS consumption and several metabolic-syndrome-related clinical parameters (Supplementary Table 6), including increased weight and waist-to-hip ratio (measures of central obesity); higher fasting blood glucose, glycosylated

haemoglobin (HbA1C%) and glucose tolerance test (GTT, measures of impaired glucose tolerance), and elevated serum alanine aminotransferase (ALT, measure of hepatic damage that is likely to be secondary, in this context, to non-alcoholic fatty liver disease). Moreover, the levels of glycosylated haemoglobin (HbA1C%), indicative of glucose concentration over the previous 3 months, were significantly increased when comparing a subgroup of high NAS consumers (40 individuals) to non-NAS consumers (236 individuals, Fig. 4a, rank sum $P < 0.002$). This increase remained significant when corrected to body mass index (BMI) levels (rank sum $P < 0.015$). In this cohort, we characterized the 16S rRNA in 172 randomly selected individuals. Notably, we found statistically significant positive correlations between multiple taxonomic entities and NAS consumption, including the *Enterobacteriaceae* family (Pearson $r = 0.36$, FDR corrected $P < 10^{-6}$), the Deltaproteobacteria class (Pearson $r = 0.33$, FDR corrected $P < 10^{-5}$) and the Actinobacteria phylum (Pearson $r = 0.27$, FDR corrected $P < 0.0003$, Supplementary Table 7). Importantly, we did not detect statistically significant correlations between OTU abundances and BMI, suggesting that the above correlations are not due to the distinct BMI of NAS consumers.

Finally, as an initial assessment of whether the relationship between human NAS consumption and blood glucose control is causative, we followed seven healthy volunteers (5 males and 2 females, aged 28–36) who do not normally consume NAS or NAS-containing foods for 1 week. During this week, participants consumed on days 2–7 the FDA's maximal acceptable daily intake (ADI) of commercial saccharin (5 mg per kg (body weight)) as three divided daily doses equivalent to 120 mg, and were monitored by continuous glucose measurements and daily GTT (Extended Data Fig. 9a). Notably, even in this short-term 7-day exposure period, most individuals (4 out of 7) developed significantly poorer glycaemic responses 5–7 days after NAS consumption (hereafter termed 'NAS responders'), compared to their individual glycaemic response on days 1–4 (Fig. 4b, c and Extended Data Fig. 9b, $P < 0.001$). None of the three NAS non-responders featured improved glucose tolerance (Fig. 4b, d and Extended Data Fig. 9c).

The microbiome configurations of NAS responders, as assessed by 16S rRNA analysis, clustered differently from non-responders both before and after NAS consumption (Fig. 4e and Extended Data Fig. 9d, respectively). Moreover, microbiomes from non-responders featured little changes in composition during the study week, whereas pronounced compositional changes were observed in NAS responders (Fig. 4f and Extended Data Fig. 9e). To study whether this NAS-induced dysbiosis has a causal role in generating glucose intolerance, stool from before (day 1, D1) or after (day 7, D7) NAS exposure were transferred from two NAS responders and two NAS non-responders into groups of normal-chow-fed germ-free mice. Indeed, transfer of post-NAS exposure (D7) stool from NAS responders induced significant glucose intolerance in recipient germ-free mice, compared to the response noted with D1 stool transferred from the same NAS-responding individuals (Fig. 4g and Extended Data Fig. 9f, $P < 0.004$ and Extended Data Fig. 9g, h, $P < 0.02$). In contrast, D7 stools transferred into germ-free mice from the two NAS non-responders induced normal glucose tolerance, which was indistinguishable from that of mice transferred with D1 stools from the same 'non-responding' individuals (Fig. 4h and Extended Data Fig. 9i–k). Germ-free mice transplanted with 'responders' microbiome replicated some of the donor saccharin-induced dysbiosis, including 20-fold relative increase of *Bacteroides fragilis* (order Bacteroidales) and *Weissella cibaria* (order Lactobacillales), and approximately tenfold decrease in *Candidatus* Arthromitus (order Clostridiales) (Extended Data Fig. 9l).

Discussion

In summary, our results suggest that NAS consumption in both mice and humans enhances the risk of glucose intolerance and that these adverse metabolic effects are mediated by modulation of the composition and function of the microbiota. Notably, several of the bacterial taxa that changed following NAS consumption were previously associated with type 2 diabetes in humans^{13,20}, including over-representation

of *Bacteroides* and under-representation of Clostridiales. Both Gram-positive and Gram-negative taxa contributed to the NAS-induced phenotype (Fig. 1a, b) and were enriched for glycan degradation pathways (Extended Data Fig. 6), previously linked to enhanced energy harvest (Fig. 2c, d)^{11,24}. This suggests that elaborate inter-species microbial cooperation may functionally orchestrate the gut ecosystem and contribute to vital community activities in diverging environmental conditions (for example, normal-chow versus high-fat dietary conditions). In addition, we show that metagenomes of saccharin-consuming mice are enriched with multiple additional pathways previously shown to associate with diabetes mellitus²³ or obesity¹¹ in mice and humans, including sphingolipid metabolism and lipopolysaccharide biosynthesis²⁵.

Our results from short- and long-term human NAS consumer cohorts (Fig. 4, Extended Data Fig. 9 and Supplementary Tables 6, 7) suggest that human individuals feature a personalized response to NAS, possibly stemming from differences in their microbiota composition and function. The changes noted in our studies may be further substantiated in mice consuming different human diets²⁶. Similarly, we believe that other individualized nutritional responses may be driven by personalized functional differences in the microbiome. As such, 'personalized nutrition' leading to 'personalized medical outcome' may underlie the variable nutritional effects noted in many multi-factorial diseases, and warrants further studies.

Artificial sweeteners were extensively introduced into our diets with the intention of reducing caloric intake and normalizing blood glucose levels without compromising the human 'sweet tooth'. Together with other major shifts that occurred in human nutrition, this increase in NAS consumption coincides with the dramatic increase in the obesity and diabetes epidemics. Our findings suggest that NAS may have directly contributed to enhancing the exact epidemic that they themselves were intended to fight. Moreover, our results point towards the need to develop new nutritional strategies tailored to the individual while integrating personalized differences in the composition and function of the gut microbiota.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 March; accepted 28 August 2014.

Published online 17 September 2014.

- Gardner, C. *et al.* Nonnutritive sweeteners: current use and health perspectives. *Diabetes Care* **35**, 1798–1808 (2012).
- Fitch, C. & Keim, K. S. Position of the Academy of Nutrition and Dietetics: use of nutritive and nonnutritive sweeteners. *Journal of the Academy of Nutrition and Dietetics* **112**, 739–758 (2012).
- Tordoff, M. G. & Alleva, A. M. Effect of drinking soda sweetened with aspartame or high-fructose corn syrup on food intake and body weight. *Am. J. Clin. Nutr.* **51**, 963–969 (1990).
- Horwitz, D. L., McLane, M. & Kobe, P. Response to single dose of aspartame or saccharin by NIDDM patients. *Diabetes Care* **11**, 230–234 (1988).
- Nettleton, J. A. *et al.* Diet soda intake and risk of incident metabolic syndrome and type 2 diabetes in the Multi-Ethnic Study of Atherosclerosis (MESA). *Diabetes Care* **32**, 688–694 (2009).
- Roberts, A., Renwick, A. G., Sims, J. & Snodin, D. J. Sucralose metabolism and pharmacokinetics in man. *Food Chem. Toxicol.* **38** (Suppl. 2), 31–41 (2000).
- Byard, J. L. & Goldberg, L. The metabolism of saccharin in laboratory animals. *Food Cosmet. Toxicol.* **11**, 391–402 (1973).
- Clemente, J. C., Ursell, L. K., Parfrey, L. W. & Knight, R. The impact of the gut microbiota on human health: an integrative view. *Cell* **148**, 1258–1270 (2012).
- Claessen, M. J. *et al.* Gut microbiota composition correlates with diet and health in the elderly. *Nature* **488**, 178–184 (2012).
- Muegge, B. D. *et al.* Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* **332**, 970–974 (2011).
- Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031 (2006).
- Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Microbial ecology: human gut microbes associated with obesity. *Nature* **444**, 1022–1023 (2006).
- Qin, J. *et al.* A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490**, 55–60 (2012).
- Hena-Mejia, J. *et al.* Inflammation-mediated dysbiosis regulates progression of NAFLD and obesity. *Nature* **482**, 179–185 (2012).
- David, L. A. *et al.* Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559–563 (2014).

16. Peterson, J. *et al.* The NIH human microbiome project. *Genome Res.* **19**, 2317–2323 (2009).
17. Koropatkin, N. M., Cameron, E. A. & Martens, E. C. How glycan metabolism shapes the human gut microbiota. *Nature Rev. Microbiol.* **10**, 323–335 (2012).
18. Schwartz, A. *et al.* Microbiota and SCFA in lean and overweight healthy subjects. *Obesity* **18**, 190–195 (2010).
19. Bergman, E. N. Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. *Physiol. Rev.* **70**, 567–590 (1990).
20. Karlsson, F. H. *et al.* Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature* **498**, 99–103 (2013).
21. Connor, S. C., Hansen, M. K., Corner, A., Smith, R. F. & Ryan, T. E. Integration of metabolomics and transcriptomics data to aid biomarker discovery in type 2 diabetes. *Mol. Biosyst.* **6**, 909–921 (2010).
22. Cani, P. D. *et al.* Changes in gut microbiota control metabolic endotoxemia-induced inflammation in high-fat diet-induced obesity and diabetes in mice. *Diabetes* **57**, 1470–1481 (2008).
23. Markle, J. G. *et al.* Sex differences in the gut microbiome drive hormone-dependent regulation of autoimmunity. *Science* **339**, 1084–1088 (2013).
24. Sonnenburg, J. L. *et al.* Glycan foraging *in vivo* by an intestine-adapted bacterial symbiont. *Science* **307**, 1955–1959 (2005).
25. Cani, P. D. *et al.* Metabolic endotoxemia initiates obesity and insulin resistance. *Diabetes* **56**, 1761–1772 (2007).
26. Smith, M. I. *et al.* Gut microbiomes of Malawian twin pairs discordant for kwashiorkor. *Science* **339**, 548–554 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the members of the Elinav and Segal laboratories for discussions. We acknowledge C. Bar-Nathan for germ-free mouse caretaking. We thank the Weizmann Institute management and the Nancy and Stephen Grand Israel

National Center for Personalized Medicine (INCPM) for providing financial and infrastructure support. We thank G. Malka, N. Kosower and R. Bikovsky for coordinating the human clinical trials, and M. Pevsner-Fischer, T. Avnit-Sagi and M. Lotan-Pompan for assistance with microbiome sample processing. C.A.T. is the recipient of a Boehringer Ingelheim Fonds PhD Fellowship. G.Z.-S. is supported by the Morris Kahn Fellowships for Systems Biology. This work was supported by grants from the National Institute of Health (NIH) and the European Research Council (ERC) to E.S., and support and grants to E.E. provided by Y. and R. Ungar, the Abisch Frenkel Foundation for the Promotion of Life Sciences, the Gurwin Family Fund for Scientific Research, Leona M. and Harry B. Helmsley Charitable Trust, Crown Endowment Fund for Immunological Research, estate of J. Gitlitz, estate of L. Hershkovich, Rising Tide foundation, Minerva Stiftung foundation, and the European Research Council. E.E. is the incumbent of the Rina Gudinski Career Development Chair.

Author Contributions J.S. conceived the project, designed and performed experiments, interpreted the results, and wrote the manuscript. T.K., D.Z. and G.Z.-S. performed the computational and metagenomic microbiota analysis and the analysis of the retrospective and prospective human study, and are listed alphabetically. C.A.T., O.M., A.W. and H.S. helped with experiments. Y.K. helped with the metabolic cage experiments. S.G. designed the metagenomic library protocols and generated the libraries. I.K.-G. performed the SCFA quantification experiments. D.I., N.Z., and Z.H. performed and supervised human experimentation. A.H. supervised the germ-free mouse experiments. E.S. and E.E. conceived and directed the project, designed experiments, interpreted the results, and wrote the manuscript.

Author Information Sequencing data are deposited in the European Nucleotide Archive accession PRJEB6996. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.S. (eran.segal@weizmann.ac.il) or E.E. (eran.elinav@weizmann.ac.il).

METHODS

Mice. C57Bl/6 WT adult male mice were randomly assigned (without blinding) to treatment groups and were given commercial artificial sweeteners (saccharin-, sucralose- or aspartame-based) or pure saccharin (Sigma Aldrich) in drinking water and fed a high-fat (HFD D12492, 60% kcal from fat, Research Diets) or standard polysaccharide normal-chow diet (Harlan-Teklad). Compared groups were always fed from the same batch of diet. For antibiotic treatment, mice were given a combination of ciprofloxacin (0.2 g l^{-1}) and metronidazole (1 g l^{-1}) or vancomycin (0.5 g l^{-1}) in their drinking water. All antibiotics were obtained from Sigma Aldrich. Adult male outbred Swiss-Webster mice (a widely used mouse strain in germ-free experiments) served as recipients for faecal transplants and were housed in sterile isolators (Park Bioservices). For faecal transplantation experiments, 200 mg of stool (from mouse pellets or human swabs) was resuspended in 5 ml of PBS under anaerobic conditions, vortexed for 3 min and allowed to settle by gravity for 2 min. Transplant into recipient mice were achieved by gavage with 200 μl of the supernatant and maintained on standard normal-chow diet and water throughout the experiment. All animal studies were approved by the Weizmann Institute of Science Institutional Animal Care and Usage committee (IACUC), application numbers 08680114-3 and 00550113-3; all animal experiments involving transfer of human microbiota into mice were approved by the Weizmann Institute of Science Bioethics and Embryonic Stem Cell Research oversight (ESCRO) committee.

Artificial and caloric sweeteners. The following commercially available NAS were dissolved in mice drinking water to obtain a 10% solution: Sucrazit (5% saccharin, 95% glucose), Sucralite (5% Sucralose), Sweet'n Low Gold (4% Aspartame), 10% glucose (J. T. Baker) and 10% sucrose (Sigma Aldrich) solutions were used for controls. The administered doses of 10% commercial NAS dissolved in water were well below their reported toxic dose (6.3 g per kg (body weight)²⁷, 16 g per kg (body weight)²⁸, and 4 g per kg (body weight)²⁹, for saccharin, sucralose and aspartame, respectively). For experiments conducted with pure saccharin (Sigma Aldrich) a 0.1 mg ml^{-1} solution was used in order to meet with FDA defined ADI for saccharin in humans (5 mg per kg (body weight)), according to the following calculation:

$$\frac{\text{ADI } 5 \text{ mg kg}^{-1} \text{ day}^{-1} \times \text{average mouse weight } 0.03 \text{ kg}}{\text{Average daily liquid intake } 2 \text{ ml}} = 0.075 \text{ mg ml}^{-1} \rightarrow 0.1 \text{ mg ml}^{-1}$$

Glucose and insulin tolerance tests. Mice were fasted for 6 h during the light phase, with free access to water. In all groups of mice where the drinking regime was other than water, it was substituted for water for the period of the fasting and glucose or insulin tolerance test. Blood from the tail vein was used to measure glucose levels using a glucometer (Bayer) immediately before and 15, 30, 60, 90 and 120 min after oral feeding with 40 mg glucose (J. T. Baker) or intra-peritoneal injection with 0.1 U per kg (body weight) Insulin (Biological Industries). Plasma fasting insulin levels were measured in sera collected immediately before the start of GTT using ELISA (Ultra Sensitive Mouse Insulin ELISA Kit, Crystal Chem).

Metabolic studies. Food and drink intake and energy expenditure were measured using the PhenoMaster system (TSE-Systems, Bad Homburg, Germany), which consists of a combination of sensitive feeding sensors for automated measurement and a photobeam-based activity monitoring system detects and records ambulatory movements, including rearing and climbing, in each cage. All parameters were measured continuously and simultaneously. Mice were trained singly-housed in identical cages before data acquisition. To calculate total caloric intake, the following values were used: Chow 3 kcal g^{-1} , sucrose $0.3938 \text{ kcal ml}^{-1}$, glucose 0.4 kcal ml^{-1} , saccharin $0.38 \text{ kcal ml}^{-1}$, sucralose $0.392 \text{ kcal ml}^{-1}$ and aspartame $0.38 \text{ kcal ml}^{-1}$.

In vitro anaerobic culturing. Pooled faecal matter from naive adult WT C57Bl/6 male mice was resuspended in 5 ml PBS in an anaerobic chamber (Coy Laboratory Products, 75% N_2 , 20% CO_2 , 5% H_2), vortexed for 3 min and allowed to settle by gravity for 2 min. 500 ml of the supernatant were added to a tube containing Chopped Meat Carbohydrate Broth, PR II (BD) and 500 ml of a 5 mg ml^{-1} saccharin solution or an equal volume of PBS. Every 3 days, 500 ml of culture were diluted to fresh medium containing saccharin or PBS. After 9 days, cultures were used for inoculation of germ-free mice.

Taxonomic microbiota analysis. Frozen faecal samples were processed for DNA isolation using the MoBio PowerSoil kit according to the manufacturer's instructions. 1 ng of the purified faecal DNA was used for PCR amplification and sequencing of the bacterial 16S rRNA gene. ~365bp Amplicons spanning the variable region 2 (V2) of the 16S rRNA gene were generated by using the following barcoded primers: Fwd 5'-AGAGTTTGTATCTGGCTCAG-3', Rev 5'-TGCTGCCTCCGTTAGG AGT-3'. The reactions were subsequently pooled in an equimolar ratio, purified (PCR clean kit, Promega), and used for Illumina MiSeq sequencing to a depth of at least 18,000 reads per sample (mean reads per sample $139,148 \pm 5264$ (s.e.m.)).

Reads were then processed using the QIIME (quantitative insights into microbial ecology) analysis pipeline as described^{30,31}, version 1.8. Paired-end joined sequences were grouped into operational taxonomic units (OTUs) using the UCLUST algorithm and the GreenGenes database³². Sequences with distance-based similarity of 97% or greater over median sequence length of 353 bp were assigned to the same OTU. Samples were grouped according to the treatment. Analysis was performed at each taxonomical level (phylum, genus and OTU level) separately. For each taxon, G test was performed between the different groups. *P* values were FDR-corrected for multiple hypothesis testing.

Shotgun pyrosequencing and sequence mapping. This was performed as previously described³³, with the following modifications: 1 μg of DNA was sheared using the Covaris 5200 system (Covaris, Inc., Woburn, MA, USA), followed by end repair, ligation to adapters, an 8-cycle PCR amplification (Kappa HiFi) and sequenced using an Illumina HiSeq to a minimal depth of 11,773,345 reads per sample (mean reads per sample $20,296,086 \pm 637,379$ (s.e.m.), read length 51 bp). Illumina sequence reads were mapped to the human microbiome reference genome database of the Human Microbiome Project (<http://hmpdacc.org/HMREFG/>, ref. 16), and to a gut microbial gene catalogue³⁴ using GEM mapper³⁵ with the following parameters:

-m 3 -s 0 -q offset-33 -gem-quality-threshold 26

Microbial species abundance was measured as the fraction of reads that mapped to a single species in the database. An expectation-maximization (EM) algorithm adapted from Pathoscope³⁶ was employed to determine the correct assignment of reads that mapped to more than one species. We considered only species for which at least 10% of the genome was covered (each coverage bin was 10,000-bp long) in at least one of the growth conditions (saccharin, water, or glucose). Reads mapped to the gut microbial gene catalogue were assigned a KEGG ID according to the mapping available with the catalogue. Genes were subsequently mapped to KEGG pathways, and only pathways whose gene coverage was above 0.2 were included. To calculate the contribution of different bacteria to the overrepresentation of glycan degradation pathways, reads that were mapped to genes in the gut microbial gene catalogue that belong to glycan degrading pathways were extracted and re-mapped the HMP reference genome database, seeking germs that had the highest contribution.

Short chain fatty acid quantification. To determine the level of free fatty acids analytic HPLC (Agilent 1260) was performed as described previously³⁷. In brief, standard solutions of acetate, butyrate and propionate (all from Sigma-Aldrich) were prepared at various concentrations (0.01–0.2 M). These solutions were analysed using HPLC, successive with quadrupole time-of-flight mass spectrometry with a step-gradient of solvent solution from 0% to 60% of CH_3CN with 0.1% formic acid to obtain calibration curve for each fatty acid. Faecal media samples were dissolved with 0.1% formic acid and analysed in a similar manner to measure the total concentration of all three free fatty acids.

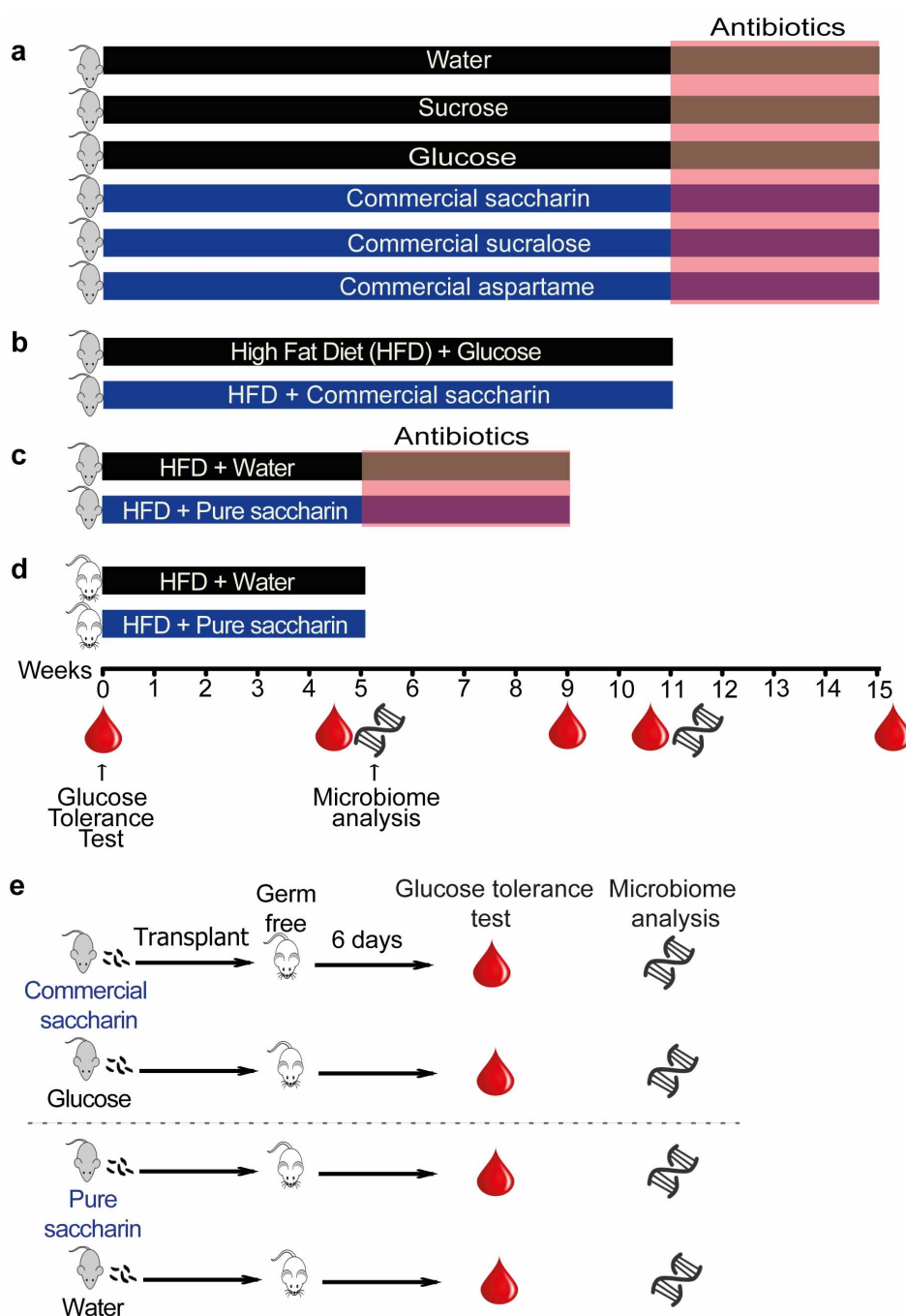
Analysis of the relationship between NAS consumption and clinical parameters in humans. All human studies were approved by the Tel Aviv Sourasky Medical Center Institutional Review Board, approval number TLV-0658-12, TLV-0050-13 and TLV-0522-10; Kfar Shaul Hospital Institutional Review Board, approval number 0-73; and the Weizmann Institute of Science Bioethics and Embryonic Stem Cell Research oversight (ESCRO) committee. The trial was reported to <http://clinicaltrials.gov/>, identifier NCT01892956. The study did not necessitate or involve randomization. For each individual in the clinical nutritional study, after signing an informed consent, multiple parameters were collected including BMI, body circumferences, fasting glucose levels, general questionnaire, complete blood counts and general chemistry parameters, a validated long-term food frequency questionnaire^{38–40}.

Long-term NAS consumption was quantified directly from answers to an explicit question regarding artificial sweeteners that participants filled out in their food frequency questionnaire. We then used the Spearman correlation to examine the relationship between NAS consumption and each of the above parameters, and FDR corrected for the multiple hypotheses tests performed.

Statistics. The following statistical analyses were used: in GTT, a two-way ANOVA and Bonferroni post-hoc analysis were used to compare between groups in different time-points, and one-way ANOVA and Tukey post hoc analysis or unpaired two-sided Student *t*-test were used to compare between AUC of multiple or two groups, respectively. Bartlett's or F-test for equal variance were employed and no significant difference was observed between variances of the compared groups. For comparison of taxonomic data, a G-test was used and *P* values were FDR-corrected for multiple hypothesis testing. In metagenomics and clinical and taxonomic data from humans, Pearson and Spearman were used for correlation tests, and Mann-Whitney U was used to compare clinical parameters between groups. *P* < 0.05 was considered

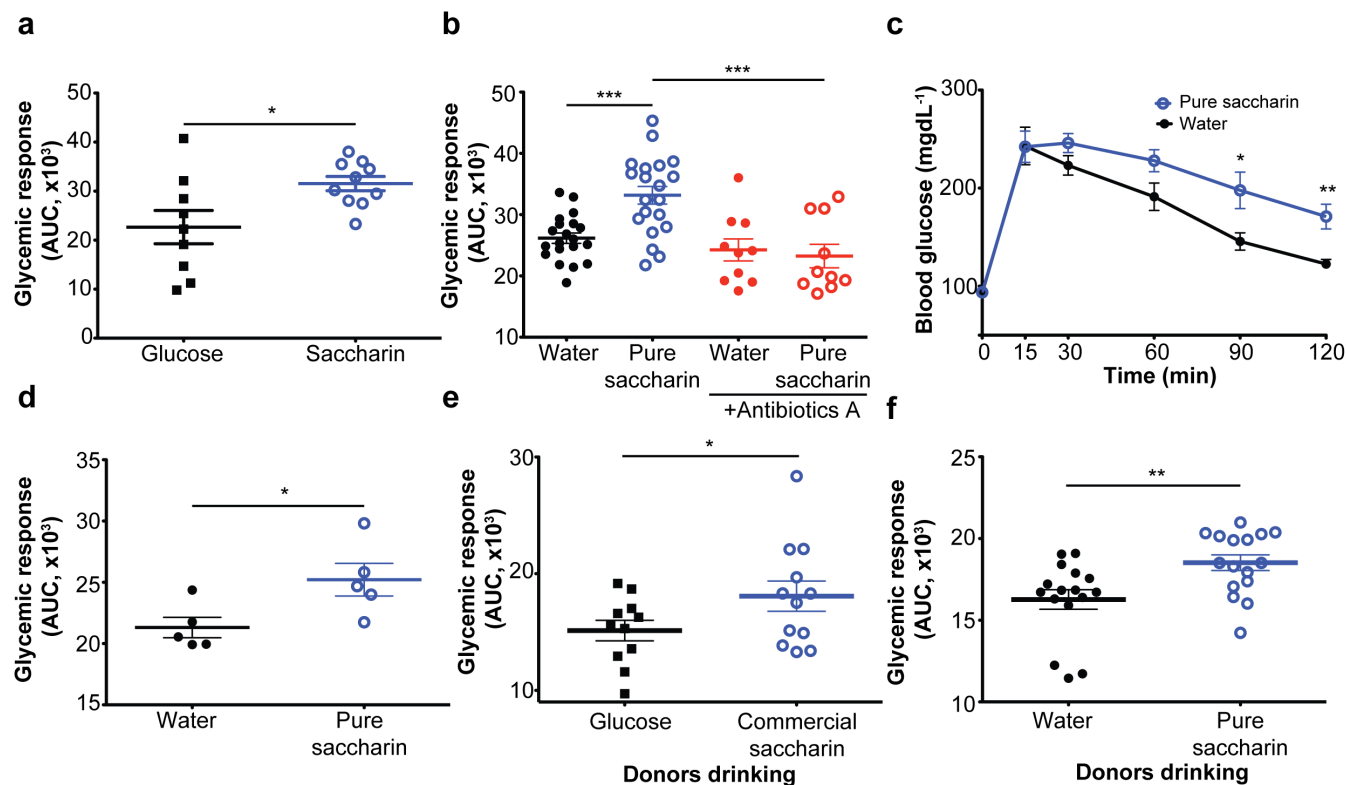
significant in all analyses (* denotes $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). In all relevant panels, symbols or horizontal lines represent the mean, and error bars represent s.e.m. For mouse experiments, cohort sizes match common practice of the described experiments. For human experiments, sample size was chosen to validate statistical analyses. No mice or data points were excluded from analyses. In the human studies, all humans older than 18 years of age who enrolled were included. Exclusion criteria included pregnancy.

27. Taylor, J. D., Richards, R. K. & Wiegand, R. G. Toxicological studies with sodium cyclamate and saccharin. *Food Cosmet. Toxicol.* **6**, 313–327 (1968).
28. Goldsmith, L. A. Acute and subchronic toxicity of sucralose. *Food Chem. Toxicol.* **38** (Suppl. 2), 53–69 (2000).
29. Magnuson, B. A. *et al.* Aspartame: a safety evaluation based on current use levels, regulations, and toxicological and epidemiological studies. *Crit. Rev. Toxicol.* **37**, 629–727 (2007).
30. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**, 335–336 (2010).
31. Elinav, E. *et al.* NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. *Cell* **145**, 745–757 (2011).
32. DeSantis, T. Z. *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72**, 5069–5072 (2006).
33. Blecher-Gonen, R. *et al.* High-throughput chromatin immunoprecipitation for genome-wide mapping of *in vivo* protein–DNA interactions and epigenomic states. *Nature Protocols* **8**, 539–554 (2013).
34. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
35. Marco-Sola, S., Sammeth, M., Guigó, R. & Ribeca, P. The GEM mapper: fast, accurate and versatile alignment by filtration. *Nature Methods* **9**, 1185–1188 (2012).
36. Francis, O. E. *et al.* Pathoscope: species identification and strain attribution with unassembled sequencing data. *Genome Res.* **23**, 1721–1729 (2013).
37. Kolodkin-Gal, I. *et al.* D-amino acids trigger biofilm disassembly. *Science* **328**, 627–629 (2010).
38. Shahar, D., Fraser, D., Shai, I. & Vardi, H. Development of a food frequency questionnaire (FFQ) for an elderly population based on a population survey. *J. Nutr.* **133**, 3625–3629 (2003).
39. Shahar, D., Shai, I., Vardi, H., Brenner-Azrad, A. & Fraser, D. Development of a semi-quantitative Food Frequency Questionnaire (FFQ) to assess dietary intake of multiethnic populations. *Eur. J. Epidemiol.* **18**, 855–861 (2003).
40. Shai, I. *et al.* Dietary evaluation and attenuation of relative risk: multiple comparisons between blood and urinary biomarkers, food frequency, and 24-hour recall questionnaires: the DEARR study. *J. Nutr.* **135**, 573–579 (2005).



Extended Data Figure 1 | Experimental scheme. 10-week-old C57Bl/6 male mice were treated with the following dietary regimes. **a**, Drinking commercially available non-caloric artificial sweeteners (NAS; saccharin, sucralose and aspartame) or glucose, sucrose or water as controls and fed a normal-chow (NC) diet. **b**, Drinking commercially available saccharin or glucose as control

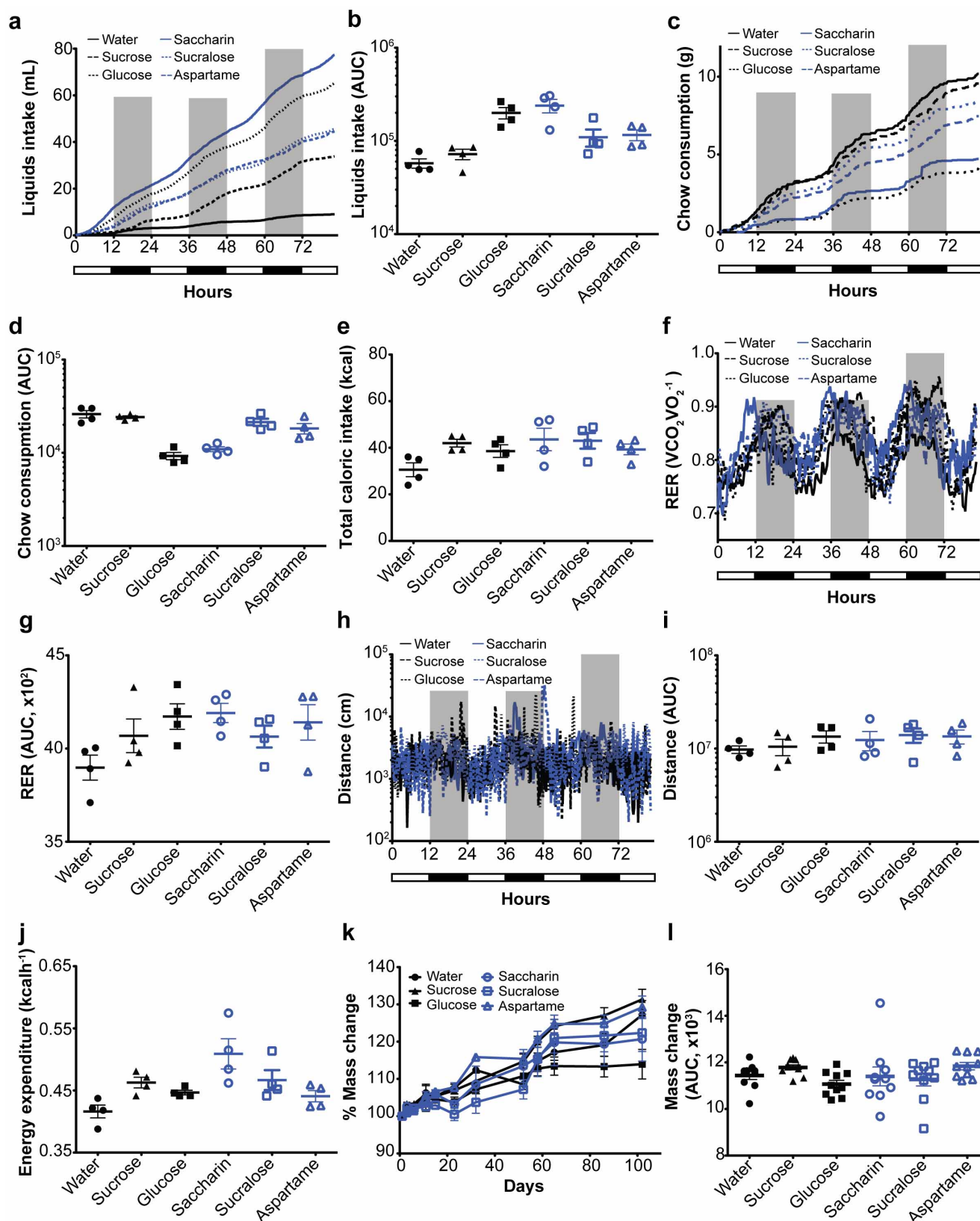
and fed a high-fat diet (HFD). **c**, Drinking pure saccharin or water and fed HFD. **d**, As in **c**, but with outbred Swiss-Webster mice. Glucose tolerance tests, microbiome analysis and supplementation of drinking water with antibiotics were performed on the indicated time points. **e**, Schematic of faecal transplant experiments.



Extended Data Figure 2 | Artificial sweeteners induce glucose intolerance.

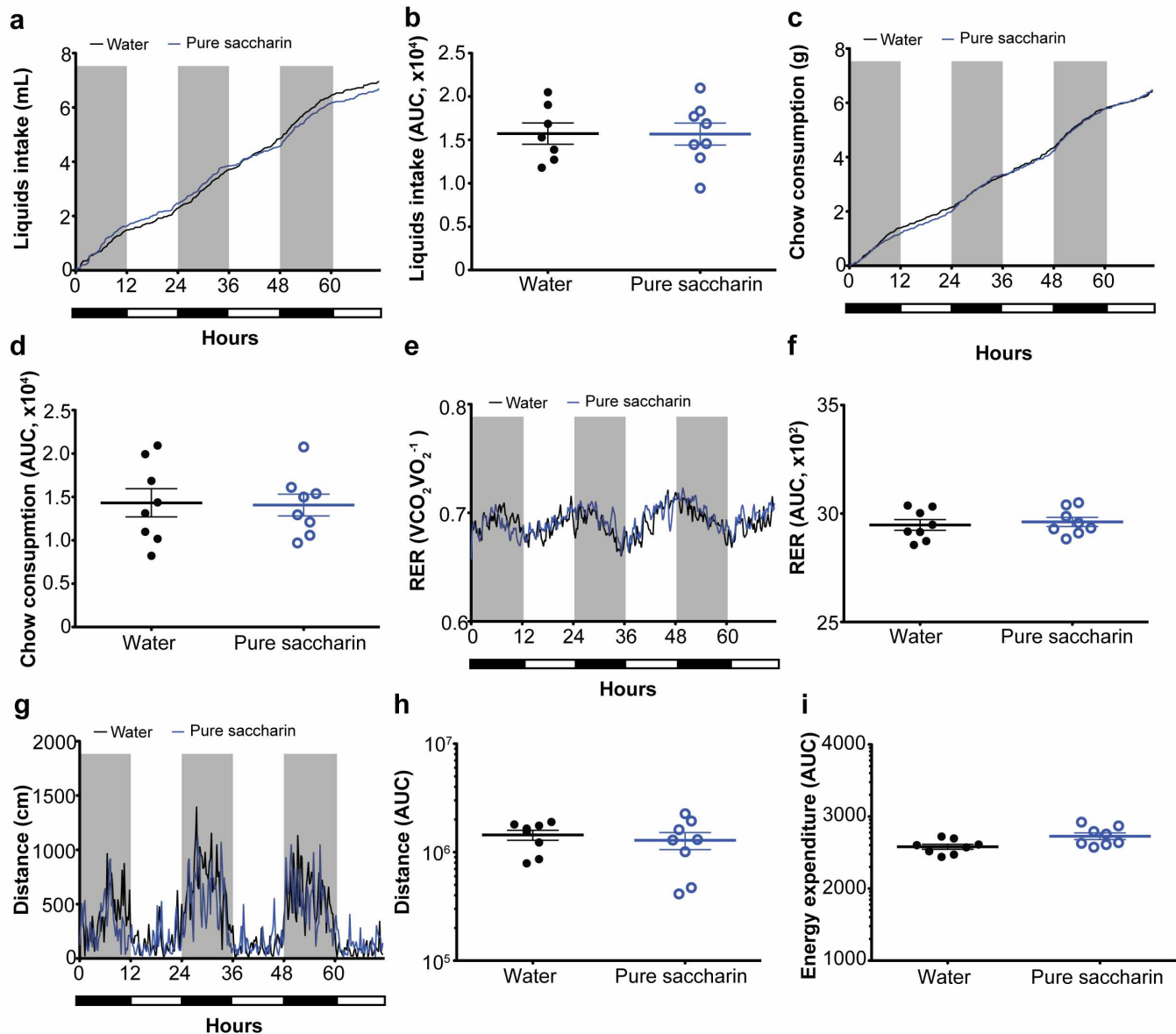
a, AUC of mice fed HFD and commercial saccharin ($N = 10$) or glucose ($N = 9$). **b**, AUC of HFD-fed mice drinking 0.1 mg ml^{-1} saccharin or water for 5 weeks ($N = 20$), followed by 'antibiotics A' ($N = 10$). **c**, **d**, OGTT and AUC of HFD-fed outbred Swiss-Webster mice ($N = 5$) drinking pure saccharin or water. **e**, **f**, Faecal samples were transferred from donor mice ($N = 10$) drinking commercially available, pure saccharin, glucose or water controls into

8-week-old male Swiss-Webster germ-free recipient mice. AUC of germ-free mice 6 days following transplant of microbiota from commercial saccharin- ($N = 12$) and glucose-fed mice ($N = 11$) (**e**); or pure saccharin- ($N = 16$) and water-fed ($N = 16$) donors (**f**). Symbols (GTT) or horizontal lines (AUC), means; error bars, s.e.m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, ANOVA and Tukey post hoc analysis (GTT) or unpaired two-sided Student *t*-test (AUC). Each experiment was repeated twice.



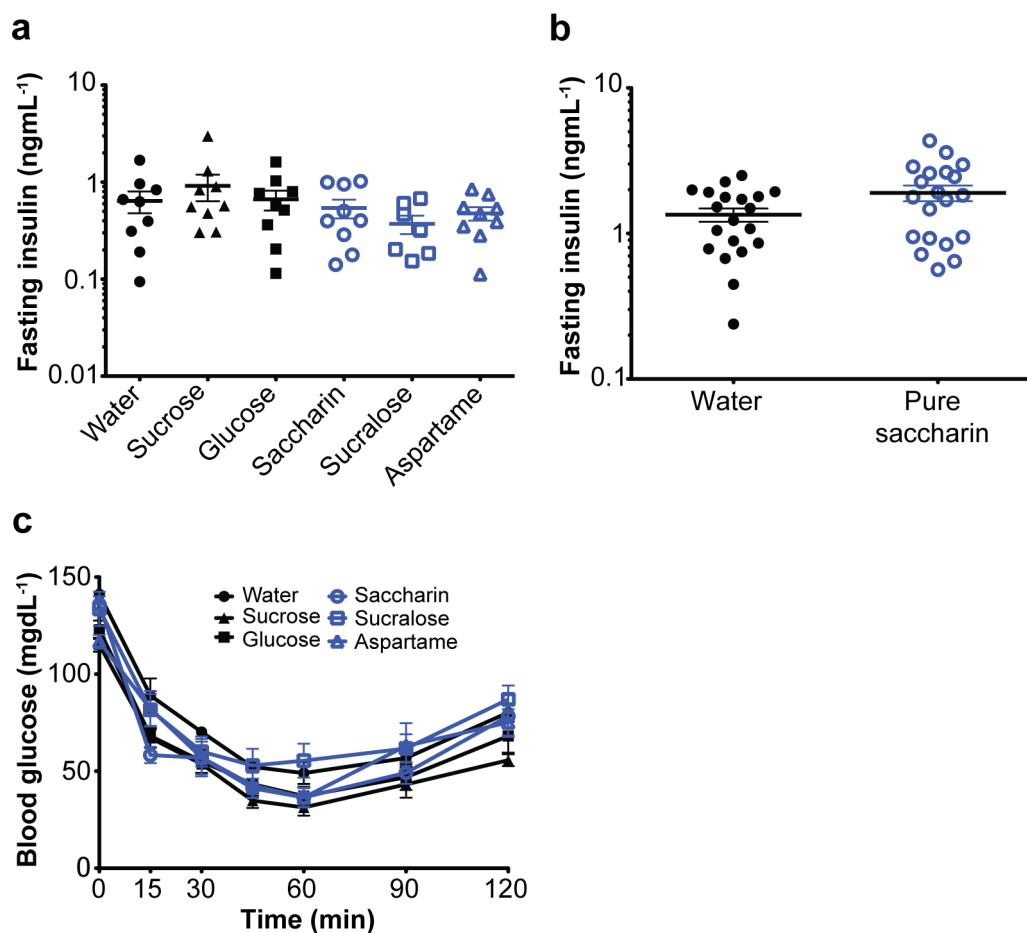
Extended Data Figure 3 | Metabolic characterization of mice consuming commercial NAS formulations. 10-week-old C57Bl/6 mice ($N = 4$) were given commercially available artificial sweeteners (saccharin, sucralose and aspartame) or controls (water, sucrose or glucose, $N = 4$ in each group) and fed normal-chow diet. After 11 weeks, metabolic parameters were characterized using the PhenoMaster metabolic cages system for 80 h. Light and dark phases are denoted by white and black rectangles on the x -axis, respectively, and grey

bars for the dark phase. **a**, Liquids intake. **b**, AUC of **a**. **c**, Chow consumption. **d**, AUC of **c**. **e**, Total caloric intake from chow and liquid during 72 h (see methods for calculation). **f**, Respiratory exchange rate (RER). **g**, AUC of **f**. **h**, Physical activity as distance. **i**, AUC of **h**. **j**, Energy expenditure. **k**, Mass change compared to original mouse weight during 15 weeks ($N = 10$). **l**, AUC of **k**. The metabolic cages characterization and weight-gain monitoring were repeated twice.



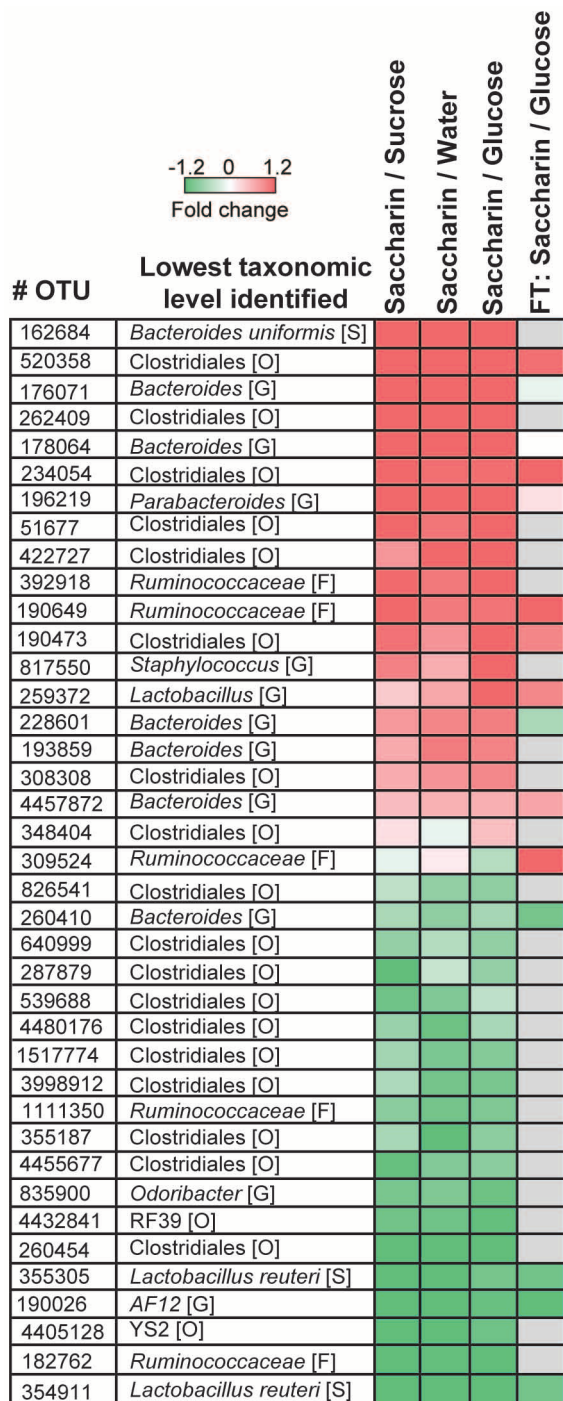
Extended Data Figure 4 | Metabolic characterization of mice consuming HFD and pure saccharin or water. 10-week-old C57Bl/6 mice ($N = 8$) were fed HFD, with or without supplementing drinking water with 0.1 mg ml^{-1} pure saccharin. After 5 weeks, metabolic parameters were characterized using the PhenoMaster metabolic cages system for 70 h. Light and dark phases are

denoted by white and black rectangles on the x -axis, respectively, and grey bars for the dark phase. **a**, Liquids intake. **b**, AUC of **a**. **c**, Chow consumption. **d**, AUC of **c**. **e**, Respiratory exchange rate (RER). **f**, AUC of **e**. **g**, Physical activity as distance. **h**, AUC of **g**. **i**, Energy expenditure. The metabolic cages characterization was repeated twice.

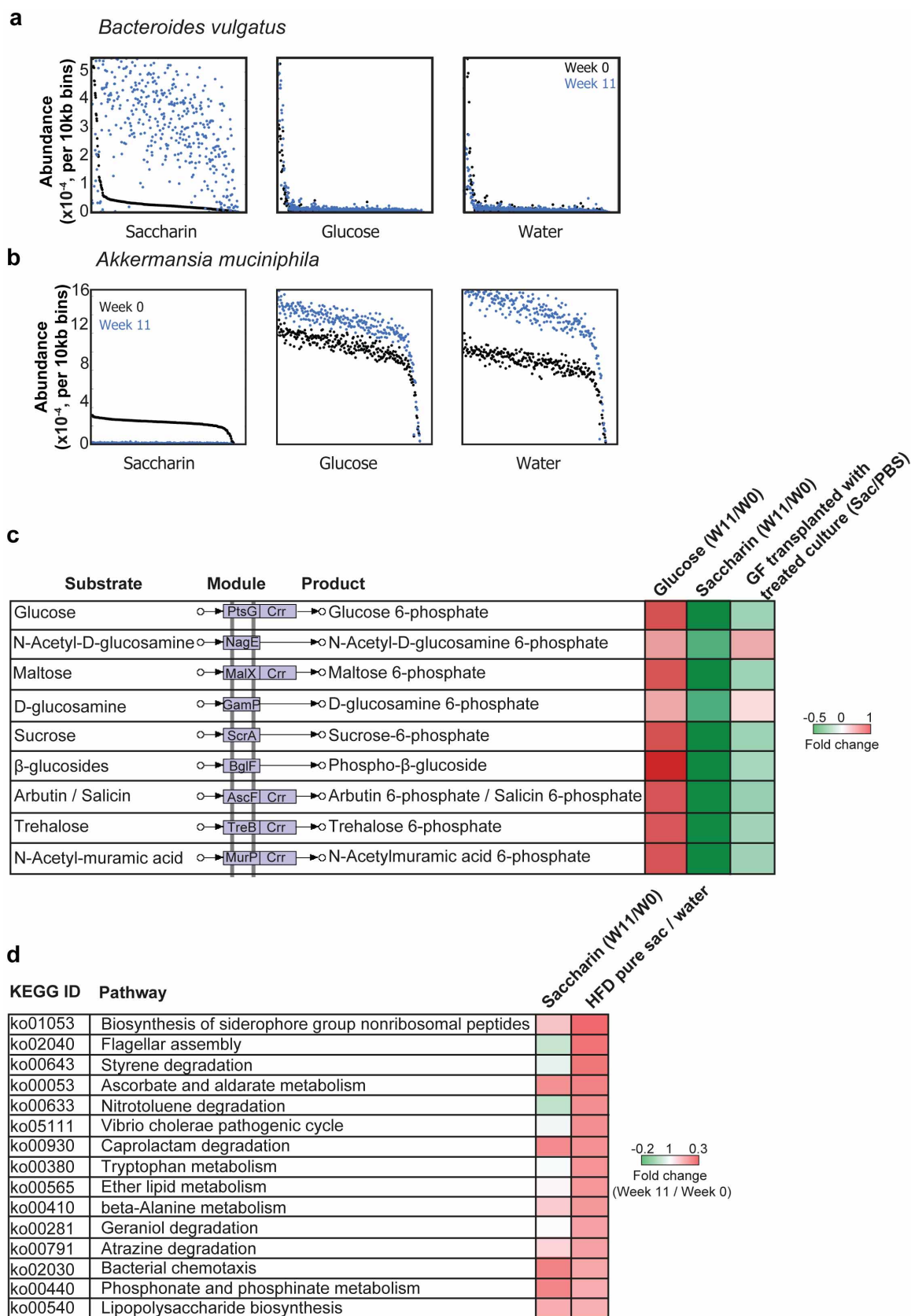


Extended Data Figure 5 | Glucose intolerant NAS-drinking mice display normal insulin levels and tolerance. **a**, Fasting plasma insulin measured after 11 weeks of commercial NAS or controls ($N = 10$). **b**, Same as **a**, but measured after 5 weeks of HFD and pure saccharin or water ($N = 20$). **c**, Insulin tolerance

test performed after 12 weeks of commercial NAS or controls ($N = 10$). Horizontal lines (**a**, **b**) or symbols (**c**) represent means; error bars, s.e.m. All measurements were performed on two independent cohorts.

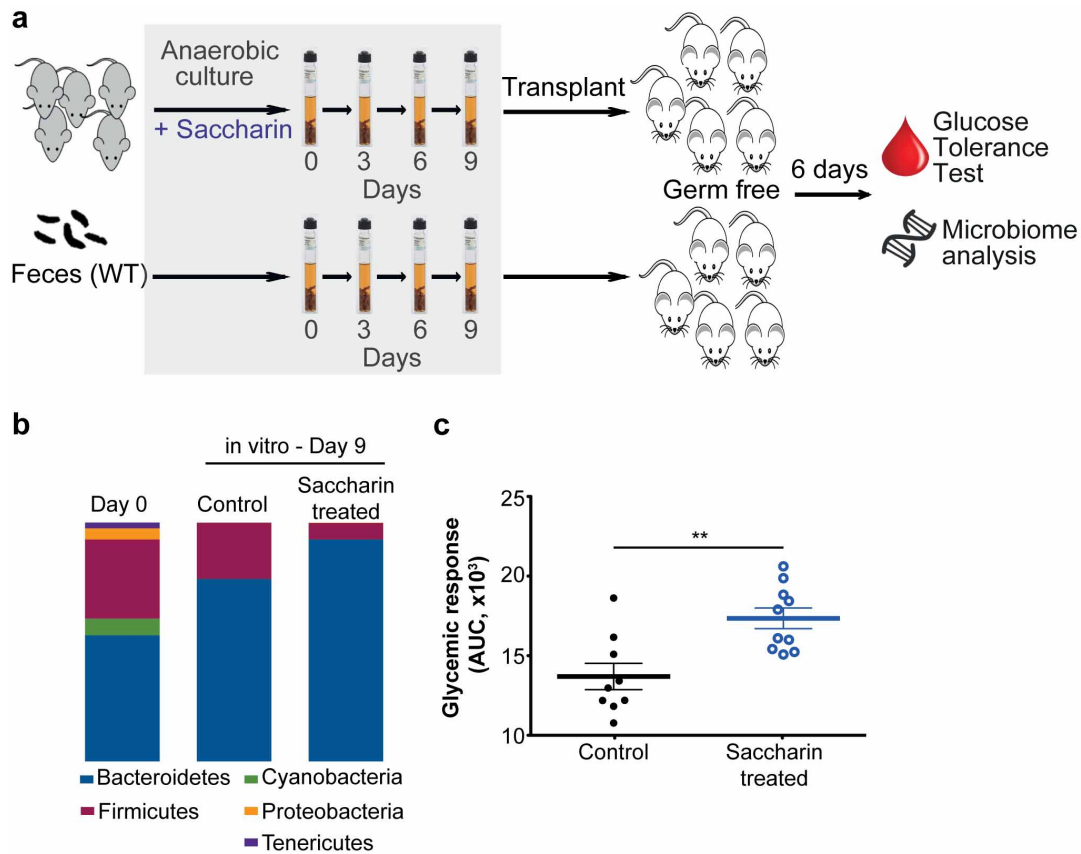


Extended Data Figure 6 | Dysbiosis in saccharin-consuming mice and germ-free recipients. Heat map representing W11 logarithmic-scale fold taxonomic differences between commercial saccharin and water or caloric sweetener consumers ($N = 5$ in each group). Right column, taxonomical differences in germ-free mice following faecal transplantation from commercial saccharin- (recipients $N = 15$) or glucose-consuming mice ($N = 13$). OTU number (GreenGenes) and the lowest taxonomic level identified are denoted.



Extended Data Figure 7 | Functional analysis of saccharin-modulated microbiota. **a, b,** Changes in bacterial relative abundance occur throughout the bacterial genome. Shown are changes in sequencing coverage along 10,000 bp genomic regions of *Bacteroides vulgatus* (**a**) and *Akkermansia muciniphila* (**b**), with bins ordered by abundance in week 0 of saccharin-treated mice. **c,** Fold change in relative abundance of modules belonging to phosphotransferase

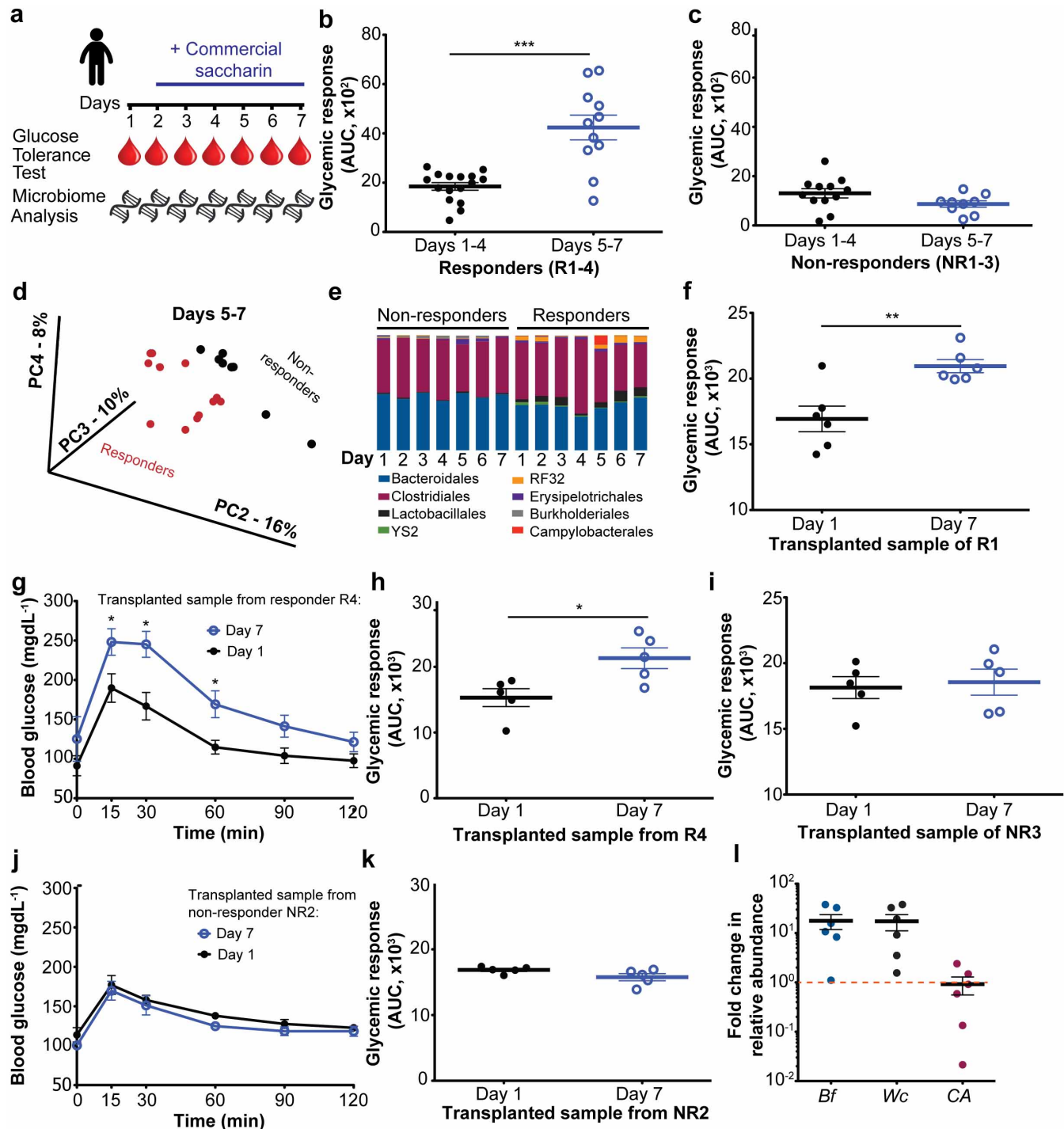
systems (PTS) between week 11 and week 0 in mice drinking commercial saccharin, glucose or water. Module diagram source: KEGG database. **d,** Enriched KEGG pathways (fold change ≥ 1.38 as cutoff) in mice consuming HFD and pure saccharin versus water compared to the fold change in relative abundance of the same pathways in mice consuming commercial saccharin (week 11/week 0).



Extended Data Figure 8 | Saccharin directly modulates the microbiota.

a, Experimental schematic. **b**, Relative taxonomic abundance of anaerobically cultured microbiota. **c**, AUC of germ-free mice 6 days following transplantation

with saccharin-enriched or control faecal cultures ($N = 10$ and $N = 9$, respectively). Horizontal lines, means; error bars, s.e.m. $**P < 0.01$, unpaired two-sided Student *t*-test. The experiment was repeated twice.



Extended Data Figure 9 | Impaired glycaemic control associated with acute saccharin consumption in humans is transferable to germ-free mice.

a, Experimental schematic (N = 7). **b**, **c**, Daily incremental AUC of days 1–4 versus days 5–7 in four responders (**b**) or three non-responders (**c**). **d**, Principal coordinates analysis (PCoA) of weighted UniFrac distances of 16S rRNA sequences demonstrating separation on principal coordinates 2 (PC2), 3 (PC3) and 4 (PC4) of microbiota from responders (N samples = 12) versus non-responders (N = 8) during days 5–7. **e**, Order-level relative abundance of taxa samples from days 1–7 of responders and non-responders. **f**, AUC in germ-free mice (N = 6) 6 days following faecal transplantation from samples of responder 1 (R1) collected before and after 7 days of saccharin consumption. **g**, **h**, OGTT and AUC in germ-free mice (N = 5) 6 days after receiving faecal samples

collected from responder 4 (R4) before and after 7 days of saccharin consumption. **i**, AUC in germ-free mice (N = 5) 6 days following faecal transplantation from samples of non-responder 3 (NR3) collected before and after 7 days of saccharin consumption. **j**, **k**, OGTT and AUC in germ-free mice (N = 5) 6 days after receiving faecal samples collected from non-responder 2 (NR2) before and after 7 days of saccharin consumption. **l**, Fold taxonomical abundance changes of selected OTUs, altered in germ-free recipients of D7 versus D1 microbiomes from R1. Dot colour same as in **e**, bacterial orders. Symbols (GTT) or horizontal lines (AUC), means; error bars, s.e.m. **P* < 0.05, ***P* < 0.01, ****P* < 0.001, two-way ANOVA and Bonferroni post-hoc analysis (GTT), unpaired two-sided Student *t*-test (AUC).

Inflammatory caspases are innate immune receptors for intracellular LPS

Jianjin Shi^{1,2*}, Yue Zhao^{2*}, Yupeng Wang², Wenqing Gao², Jingjin Ding^{2,3}, Peng Li², Liyan Hu² & Feng Shao^{1,2,3,4}

The murine caspase-11 non-canonical inflammasome responds to various bacterial infections. Caspase-11 activation-induced pyroptosis, in response to cytoplasmic lipopolysaccharide (LPS), is critical for endotoxic shock in mice. The mechanism underlying cytosolic LPS sensing and the responsible pattern recognition receptor are unknown. Here we show that human monocytes, epithelial cells and keratinocytes undergo necrosis upon cytoplasmic delivery of LPS. LPS-induced cytotoxicity was mediated by human caspase-4 that could functionally complement murine caspase-11. Human caspase-4 and the mouse homologue caspase-11 (hereafter referred to as caspase-4/11) and also human caspase-5, directly bound to LPS and lipid A with high specificity and affinity. LPS associated with endogenous caspase-11 in pyroptotic cells. Insect-cell purified caspase-4/11 underwent oligomerization upon LPS binding, resulting in activation of the caspases. Underacylated lipid IVa and lipopolysaccharide from *Rhodobacter sphaeroides* (LPS-RS) could bind to caspase-4/11 but failed to induce their oligomerization and activation. LPS binding was mediated by the CARD domain of the caspase. Binding-deficient CARD-domain point mutants did not respond to LPS with oligomerization or activation and failed to induce pyroptosis upon LPS electroporation or bacterial infections. The function of caspase-4/5/11 represents a new mode of pattern recognition in immunity and also an unprecedented means of caspase activation.

LPS, the major structural element of Gram-negative bacteria outer membrane, triggers strong immune responses. Excessive LPS resulting from uncontrolled infection causes septic shock. The caspase-11 non-canonical inflammasome in mice can sense various bacterial infections including *Escherichia coli*¹, *Salmonella typhimurium*², *Legionella pneumophila*^{3–5} and *Burkholderia thailandensis*⁵ by responding to cytoplasmic LPS^{6,7}. Caspase-11 activation plays an important role in endotoxic shock and sepsis^{6–8}. Caspase-11-mediated innate immunity is promoted by type I interferon (IFN) and IFN-induced guanylate-binding proteins that facilitate caspase-11 transcription^{9,10} and lysis of bacteria-containing vacuoles^{11,12}, respectively. However, whether an intracellular LPS-sensing pathway operates in humans and how LPS activates the non-canonical inflammasome are unknown.

Caspase-4 detects LPS in human monocytes

To dissect cytoplasmic sensing of LPS, we developed an efficient LPS transfection protocol using a commercial electroporation device. Electroporation of ultrapure *E. coli* LPS (O111:B4) into human U937 monocytes caused massive cell death (Fig. 1a, b). Electroporation of the bacterial cell wall fragment muramyl dipeptide (MDP) did not induce cell death. Human THP1 monocytes showed the same differential responses to LPS and MDP (Fig. 1a, b). LPS contains lipid A, a core polysaccharide chain and a serotype-specific O-antigenic oligosaccharide^{13,14}. LPS derived from other *E. coli* serotypes and Gram-negative bacteria as well as lipid A all induced cell death in U937 cells (Extended Data Fig. 1a). Cytoplasmic LPS-triggered monocyte death morphologically resembled caspase-1-mediated pyroptosis (Fig. 1a). However, *CASP1*^{−/−} U937 cells generated by CRISPR/Cas9-mediated targeting remained sensitive to LPS electroporation (Extended Data Fig. 1b). The ASC adaptor is often involved in inflammasome activation but short hairpin RNA (shRNA) knockdown of ASC¹⁵ showed no inhibitory effect (Extended Data Fig. 1b). LPS-induced

cytotoxicity was resistant to caspase-1 inhibitor Z-YVAD-FMK but blocked by the pan-caspase-inhibitor zVAD-FMK (Extended Data Fig. 1c).

Humans do not possess caspase-11. Human caspases-4 and -5 are most homologous to caspase-11. Caspases 4, 5 and 11 are classified as inflammatory caspases. Reverse-transcription PCR (RT-PCR) failed to detect *CASP5*, but did detect *CASP4* expression in U937 and THP1 cells (Fig. 1c). Immunoblotting confirmed the abundant expression of caspase-4 in both types of monocytes (Fig. 1c). Short interfering RNA (siRNA) knockdown of *CASP4* strongly inhibited LPS-induced pyroptosis in U937 and THP1 cells (Fig. 1d). *CASP4* knockdown did not affect pyroptosis triggered by double-stranded DNA (dsDNA) activation of the AIM2 inflammasome. The requirement of caspase-4 was further confirmed by three independent *CASP4*-specific short hairpin RNAs (shRNAs) (Extended Data Fig. 1d). Murine caspase-11 mediates flagellin-independent pyroptosis triggered by *L. pneumophila*^{3,4}, which is enhanced by deletion of a type IV secretion system effector SdhA important for *Legionella*-containing vacuole membrane integrity⁵. *L. pneumophila* Δ sdhA-infected U937 cells underwent pyroptosis, which was severely blocked by stable knockdown of *CASP4* but not ASC (Fig. 1e). Thus, caspase-4 mediates cytoplasmic LPS-induced pyroptosis in human monocytes.

Caspase-4 detects LPS in human non-monocytes

Keratinocytes and certain epithelial cells also express caspase-4 (refs 16–18). LPS electroporation caused massive cell death in HeLa, HaCaT, HT29 and HL60 cells but not in 293T and Jurkat cells (Fig. 1f). Consistent with the differential responses, caspase-4 was only expressed in HeLa, HaCaT, HT29 and HL60 cells (Fig. 1g). Similar to monocytes, these non-macrophage cells resisted MDP electroporation. siRNA knockdown of *CASP4* in HaCaT, HT29 and HeLa cells all blocked cytoplasmic LPS-induced cytotoxicity (Fig. 1h and Extended Data Fig. 1e); control siRNA and a

¹Peking University-Tsinghua University-National Institute of Biological Sciences Joint Graduate Program, National Institute of Biological Sciences, Beijing 102206, China. ²National Institute of Biological Sciences, Beijing 102206, China. ³National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China. ⁴National Institute of Biological Sciences, Beijing, Collaborative Innovation Center for Cancer Medicine, Beijing 102206, China.

*These authors contributed equally to this work.

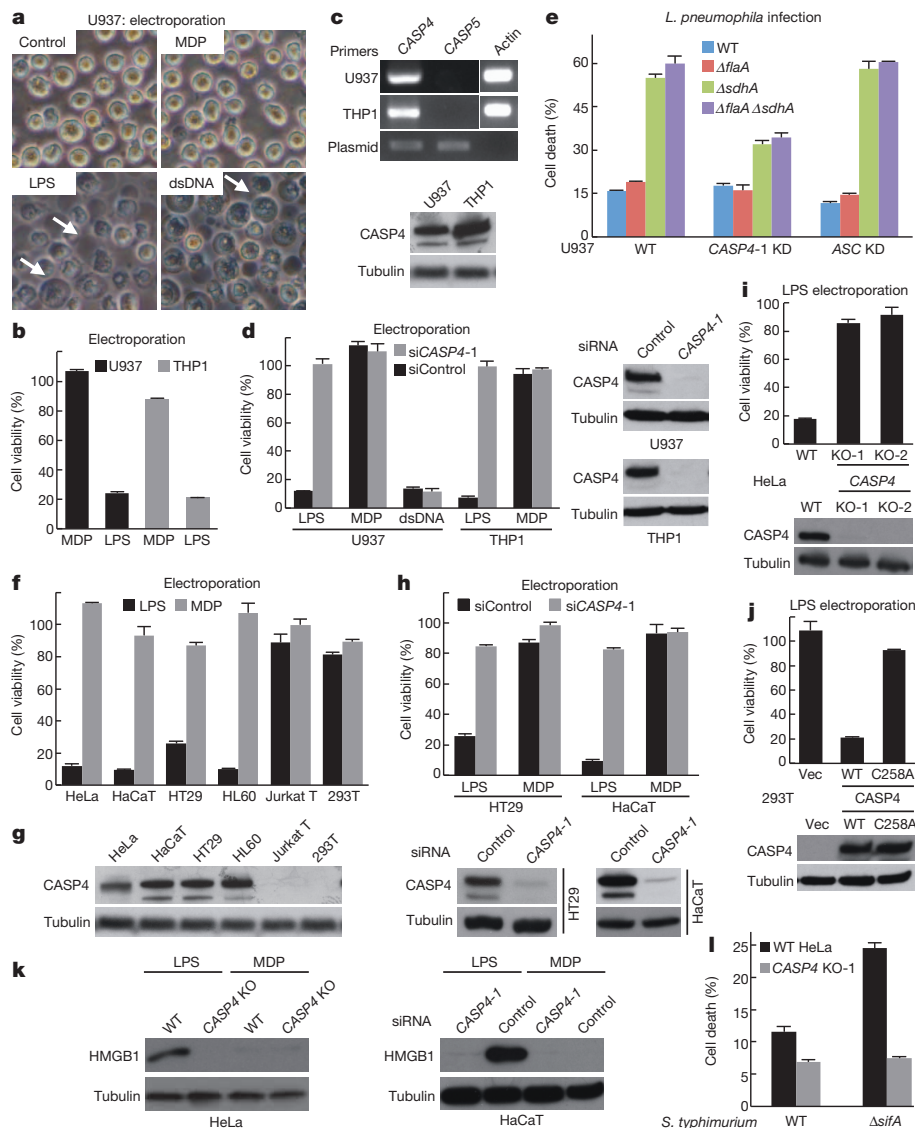


Figure 1 | Caspase-4 mediates cytoplasmic LPS-induced cytotoxicity in human macrophage and non-macrophage cells. **a**, Morphology of U937 cells upon MDP, LPS or dsDNA electroporation. Arrows, pyroptotic cells. Control, electroporation alone. **b**, Viability of U937 and THP1 cells upon MDP and LPS electroporation. **c**, Caspase-4/5 expression in U937 and THP1 cells. For the upper RT-PCR assay, plasmid-borne caspase-4/5 serves as the control templates. Immunoblotting of total lysates using indicated antibodies as shown below. **d, h, i**, Effects of *CASP4* knockdown or knockout on cytoplasmic LPS-induced cytotoxicity. *CASP4*-specific (si*CASP4*-1) or a control siRNA (siControl) was electroporated (d) or chemically transfected in (h) into cells. Two independent clones of *CASP4*^{-/-} HeLa cells (KO-1/2) were assayed in i. Indicated agonists were electroporated into cells. Anti-caspase-4 immunoblots show the knockdown/knockout efficiency. **e**, Effects of *CASP4* knockdown on *L. pneumophila*-induced

CASP4-specific siRNA showed no inhibition. CRISPR/Cas9-mediated knockout of *CASP4* in HeLa cells completely abolished LPS-induced cytotoxicity (Fig. 1i). Conversely, ectopic expression of caspase-4 in 293T cells rendered a high sensitivity to LPS electroporation, however, cells expressing the protease-deficient caspase-4(C258A) mutant were not sensitive to LPS treatment (Fig. 1j). Caspase-4-mediated pyroptosis was accompanied by secretion of a pro-inflammatory mediator HMGB1 (Fig. 1k). *Salmonella enterica* serovar Typhimurium (hereafter referred to as *S. typhimurium*) lacking a type III secretion system effector SifA triggers caspase-11-dependent death in mouse macrophages⁶. *S. typhimurium* Δ sifA also induced more HeLa cell death than wild-type bacteria, which was diminished in *CASP4*^{-/-} cells (Fig. 1l). The amounts of cytosolic LPS detected with *S. typhimurium* infection were approximately

six times lower than those with LPS electroporation (Extended Data Fig. 1f), and this expected phenomenon is consistent with the much lower cell death observed with bacterial infection. Thus, human non-macrophage cells responds to cytoplasmic LPS through caspase-4-mediated inflammatory cell death.

six times lower than those with LPS electroporation (Extended Data Fig. 1f), and this expected phenomenon is consistent with the much lower cell death observed with bacterial infection. Thus, human non-macrophage cells responds to cytoplasmic LPS through caspase-4-mediated inflammatory cell death.

Functional interchange of caspase-4 and -11

The high-efficiency LPS electroporation worked equally well in mouse primary and immortalized bone marrow-derived macrophages (priBMDM and iBMDM, respectively) (Extended Data Fig. 2a, b). The cell death was independent of Ripk3 and Mlkl, key mediators of necroptosis (Extended Data Fig. 2a). Knockout of *Casp11* abolished cytoplasmic LPS-induced pyroptosis in iBMDMs, which was restored by stable re-expression of

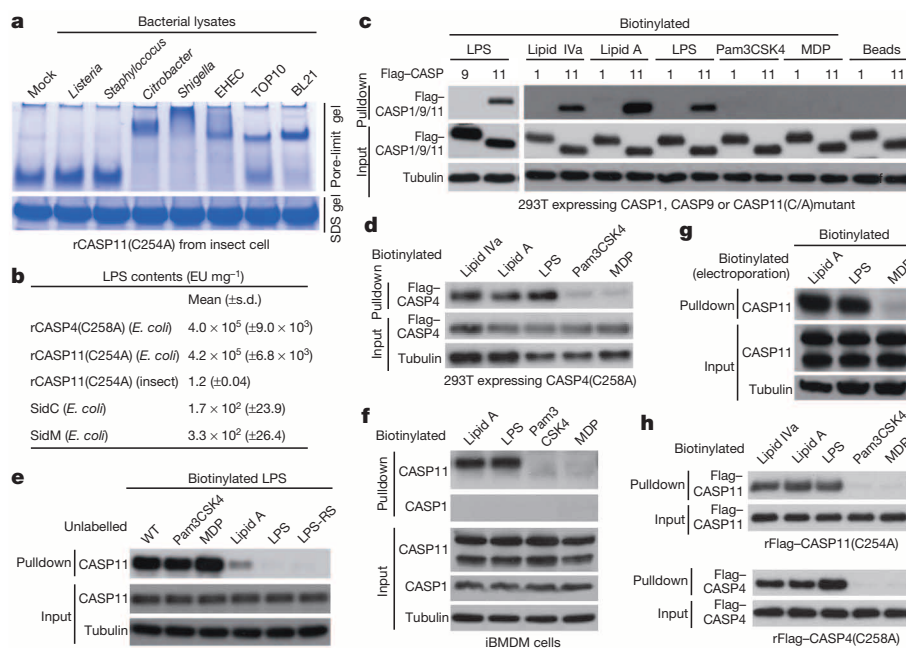


Figure 2 | Caspase-4/11 directly and specifically binds to LPS. **a**, Pore-limit native gel analysis of insect-cell purified caspase-11(C254A) mutant (rCASP11(C254A)) treated with indicated bacterial lysates. The corresponding SDS-PAGE gel is shown below. **b**, LPS contents in *E. coli* and insect-cell purified caspase-4(C258A) and caspase-11(C254A) proteins in comparison with bacterially purified SidC and SidM. The values are mean values \pm s.d. from three replicates. **c–e**, Streptavidin pull-down assays of the binding of biotin-conjugated lipid A, lipid IVa, LPS, Pam3CSK4 and MDP to Flag tagged caspase-11(C254A) and caspase-4(C258A) in transfected 293T cell lysates.

caspase-11 but not the protease-deficient caspase-11(C254A) mutant (Extended Data Fig. 2b). LPS electroporation-induced BMDM death did not require, but was sensitized by, prior IFN- γ stimulation (Extended Data Fig. 2c). Caspase-4 showed a similar complementation effect, requiring caspase-4 Cys 258 (Extended Data Fig. 2b). In a reciprocal experiment, stable expression of caspase-11 in human 293T cells resulted in extensive cell death upon LPS electroporation (Extended Data Fig. 2d). Thus, human caspase-4 functions similarly to mouse caspase-11 in innate sensing of cytoplasmic LPS.

Caspase-4/11 and also -5 directly to bind LPS

Inflammatory caspases contain an amino (N)-terminal caspase-activation and recruitment domain (CARD) and are believed to be activated by a CARD-containing protein. We examined 18 mammalian CARD-domain proteins, but none could induce caspase-11 dependent cell death in transfected 293T cells (Extended Data Fig. 3). When preparing recombinant caspase-4/11 for functional characterization, we noticed an unexpected phenomenon. Caspase-11(C254A) and caspase-4(C258A) purified from Sf21 insect cells eluted at a size of ~ 100 kDa from a gel filtration column, whereas those from *E. coli* were oligomers of ~ 600 kDa (Extended Data Fig. 4a, b). The 100-kDa caspase-4/11 were monomers according to the molecular mass determined from analytical ultracentrifugation and static light scattering (Extended Data Fig. 4c). Consistent with this result, insect-cell purified caspase-11 treated with crosslinking agents remained as monomers on the SDS gel (Extended Data Fig. 4d). The differential oligomerization of *E. coli* and insect-cell produced caspase-4/11 was supported by their different mobility on a pore-limit native gel (Extended Data Fig. 4a, b).

We speculated that some bacterial component accounts for the appearance of the caspase-4/11 oligomer. Insect-cell purified caspase-11(C254A) was incubated with lysates from several bacteria species. Although incubation with Gram-positive bacterial lysates did not affect the oligomerization state of caspase-11, lysates from *E. coli* TOP10 and

Indicated unlabelled agonists were used for competition in **e**, **f**, **g**. Streptavidin pull-down assays of LPS/lipid A binding to endogenous caspase-11. Lysates of LPS-primed iBMDMs were incubated with indicated biotinylated agonists in **f**; the agonists were electroporated into iBMDMs in **g**, **h**. Streptavidin pull-down assays of LPS, lipid A and lipid IVa binding to insect-cell purified caspase-11 (rFlag-CASP11(C254A)) and caspase-4 (rFlag-CASP4(C258A)). Shown are anti-Flag (**c**, **d**, **e**) and anti-caspase-1/11 (**f**, **g**) immunoblots of the pull-downs and total lysates (input). All data shown are representative of at least three independent experiments.

BL21 strains, *Citrobacter rodentium*, *Shigella flexneri* and enterohaemorrhagic *E. coli* all induced oligomerization (Fig. 2a). LPS is one of the most abundant components in Gram-negative bacteria. Therefore, we decided to investigate whether LPS accounts for caspase-11 oligomerization before performing activity-based fractionation. We first measured LPS content in caspase-4/11. In our experiments, bacterially purified recombinant proteins typically contained LPS at the level of several hundred EU per mg (Fig. 2b). Notably, the amounts of LPS in *E. coli* purified caspase-4/11 were higher by three orders of magnitude. Insect-cell purified caspase-11 contained little LPS (Fig. 2b). These results led to the hypothesis that caspase-4/11 may bind directly to LPS.

A series of pull-down assays were carried out to test this idea. Biotinylated LPS efficiently precipitated Flag tagged caspase-11(C254A) but not Flag tagged caspase-1(C285A) or caspase-9(C325A) from transfected 293T lysates (Fig. 2c). Caspase-11 was not precipitated by biotinylated MDP and Pam3CSK4 (another immunostimulatory peptide from bacterial lipoprotein) (Fig. 2c). Synthetic biotin-lipid A also precipitated Flag tagged caspase-11 but not caspase-1 (Fig. 2c). Similar results were obtained with Flag tagged caspase-4(C258A) (Fig. 2d). Evidence for the specific binding of caspase-11 to LPS was strengthened by efficient competition with unlabelled LPS/lipid A but not MDP or Pam3CSK4 (Fig. 2e). Endogenous caspase-11 but not caspase-1 in iBMDM lysates was also readily precipitated by LPS and lipid A but not MDP or Pam3CSK4 (Fig. 2f). Caspase-11 exists as two species of 38 kDa and 43 kDa (full-length)¹⁹, but only the latter was precipitated by LPS. Endogenous caspase-11 was also precipitated by LPS and lipid A but not MDP electroporated into iBMDMs (Fig. 2g), demonstrating a specific association during the course of cell pyroptosis. We also examined caspase-5, and found that both isoforms of caspase-5 (a and f) were robustly precipitated by LPS and lipid A but not MDP or Pam3CSK4 (Extended Data Fig. 5a). Consistent with this result, caspase-5 restored cytoplasmic LPS-induced pyroptosis in *Casp11*^{-/-} iBMDMs (Extended Data Fig. 5b). These results suggest caspase-5 has a similar function in recognizing cytoplasmic LPS.

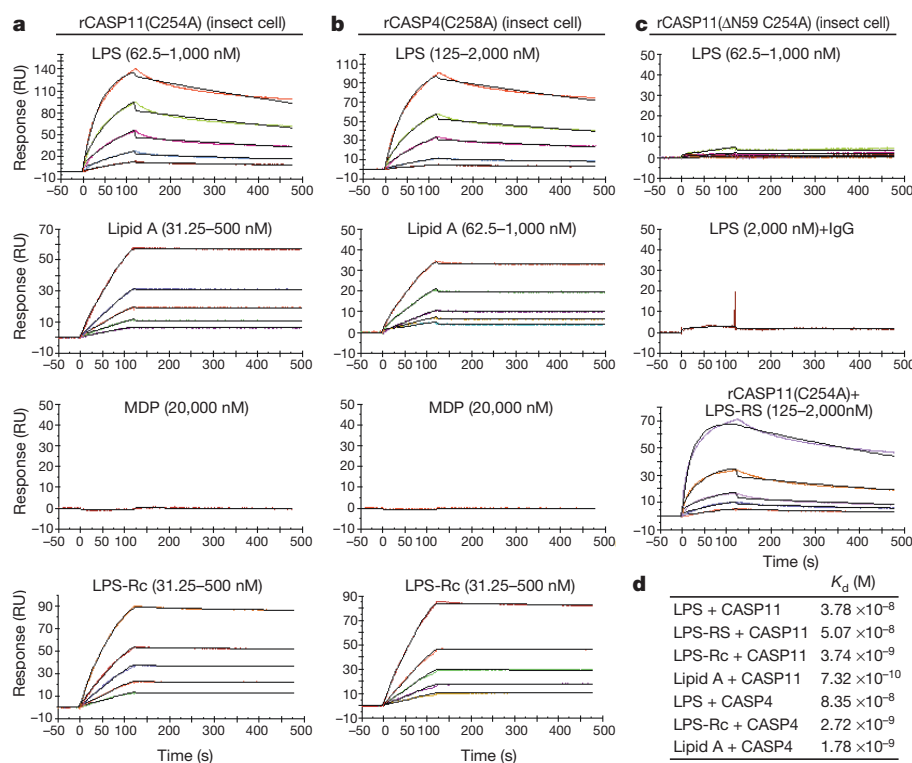


Figure 3 | Surface plasmon resonance measurements of the binding between LPS/lipid A and caspase-4/11. **a–c.** The sensorgrams of LPS (*E. coli* O111:B4) or lipid A binding to the indicated chip-immobilized proteins in **a–c** are expressed in RU (response unit) versus time after subtracting the control signal. Recombinant caspase-11 (rCASP11) and caspase-4 (rCASP4) were both in the catalytic-cysteine mutant background and purified from insect cells. The LPS concentrations were 62.5, 125, 250, 500 and 1,000 nM (from bottom to top) for caspase-11, and 125, 250, 500, 1,000 and 2,000 nM for

caspase-4. The lipid A concentrations were 31.25, 62.5, 125, 250 and 500 nM for caspase-11, and 62.5, 125, 250, 500 and 1,000 nM for caspase-4. LPS-Rc concentrations were the same as those of lipid A for caspase-11 and LPS-RS concentrations follow those of LPS for caspase-4. MDP (20,000 nM) and IgG were included as negative controls. Colour lines, SPR data from different concentrations of the analytes; black lines, model fits. **d.** The calculated dissociation constants (K_d). All data shown are representative of at least two independent experiments.

High-affinity LPS–caspase-4/11 binding

Similar to cellular caspase-4/11, insect-cell purified caspase-11(C254A) was robustly pulled down by biotin-conjugated LPS and lipid A, but not MDP or Pam3CSK4 (Fig. 2h). Similar results were obtained with insect-cell purified caspase-4 (Fig. 2h). Both LPS and lipid A showed a dose-dependent resonance signal and rapid association with chip-immobilized caspase-11(C254A) and caspase-4(C258A) on standard surface plasmon resonance (SPR) (Fig. 3a, b). No resonance signal was detected with a tenfold higher concentration of MDP. LPS showed no binding to IgG on the SPR (Fig. 3c). The calculated dissociation constants (K_d) between caspase-4/11 and LPS were 8.35×10^{-8} M and 3.78×10^{-8} M, respectively (Fig. 3d). The higher K_d of lipid A was mainly due to the lower dissociation rate constant (Fig. 3a, b, d). Caspase-4/11 binding to LPS features a slower dissociation rate constant, compared with known LPS-binding proteins such as CD14, TLR4/MD2 and LBP²⁰.

Transfected caspase-11 also appeared as 43 kDa and 38 kDa forms and the shorter one lacked ~60 residues from the N terminus (Extended Data Fig. 4e and data not shown). Consistently, LPS showed no binding to caspase-11(ΔN59, devoid of 59 residues from the N terminus) (Fig. 3c). Further supporting the requirement of the CARD domain, LPS co-eluted with full-length caspase-11 but not the ΔN59(C254A) mutant when the two proteins were incubated with LPS and analysed by gel filtration chromatography (Extended Data Fig. 4f). Compared with LPS and lipid A, LPS-Rc lacking the outer core polysaccharide and O-antigenic oligosaccharides showed an intermediate binding affinity (Fig. 3), indicating that the sugar in LPS may interfere with caspase-11 binding.

LPS induces caspase-4/11 oligomerization

Following LPS incubation, insect-cell purified caspase-4/11 (catalytic cysteine mutants) oligomerized at a size roughly the same as that of *E. coli*

purified caspase-4/11 (Fig. 4a). LPS co-eluted with the caspase-11 oligomer but not the binding-deficient caspase-11(ΔN59) monomer (Extended Data Fig. 4f). LPS-induced caspase-4/11 oligomerization was also observed on the pore-limit native gel (Fig. 4b). Consistent with the binding data, lipid A but not MDP or Pam3CSK4 stimulated caspase-4/11 oligomerization (Fig. 4a, b). Treatment with 0.5% Triton X-100 reduced LPS-induced caspase-11 oligomers into a smaller homogeneous form on the pore-limit gel (Fig. 4b). Using this assay, Flag tagged caspase-11(C254A) in 293T cell extracts was found to oligomerize in a way similar to the recombinant protein upon incubation with LPS/lipid A but not MDP or Pam3CSK4 (Fig. 4c). LPS variants with different lengths of polysaccharide core (LPS-Ra, Rc and Rd), as well as the sugar-free LPS-Re, also stimulated caspase-11 oligomerization (Fig. 4d). Consistently, these 'rough' LPS mutants triggered extensive pyroptosis when electroporated into BMDMs (Fig. 4e). Lipid A induced less cell death (Fig. 4e and Extended Data Fig. 1a) due to its lower solubility affecting the electroporation efficiency. Caspase-4/11 harbouring intact catalytic cysteine showed a similar oligomerization upon lipid A incubation, which did not occur with the ΔN59 mutant (Extended Data Fig. 4g).

We generated a caspase-1 and caspase-11 chimera, in which the CARD domain in caspase-1 was replaced with the CARD domain of caspase-11. This chimaeric protein was fully functional in sensitizing *Casp11*^{-/-} BMDMs, *CASP4*^{-/-} HeLa cells, as well as 293T cells to LPS-induced cytotoxicity (Extended Data Fig. 6a). Consistent with this data, the chimaeric protein could bind to LPS and underwent oligomerization upon incubation with LPS but not MDP (Extended Data Fig. 6b, c). The CARD domain alone, but not caspase-11(ΔCARD), was sufficient for LPS binding (Extended Data Fig. 6b). Transfected caspase-11 CARD domain alone was a monomer but formed an oligomer upon addition

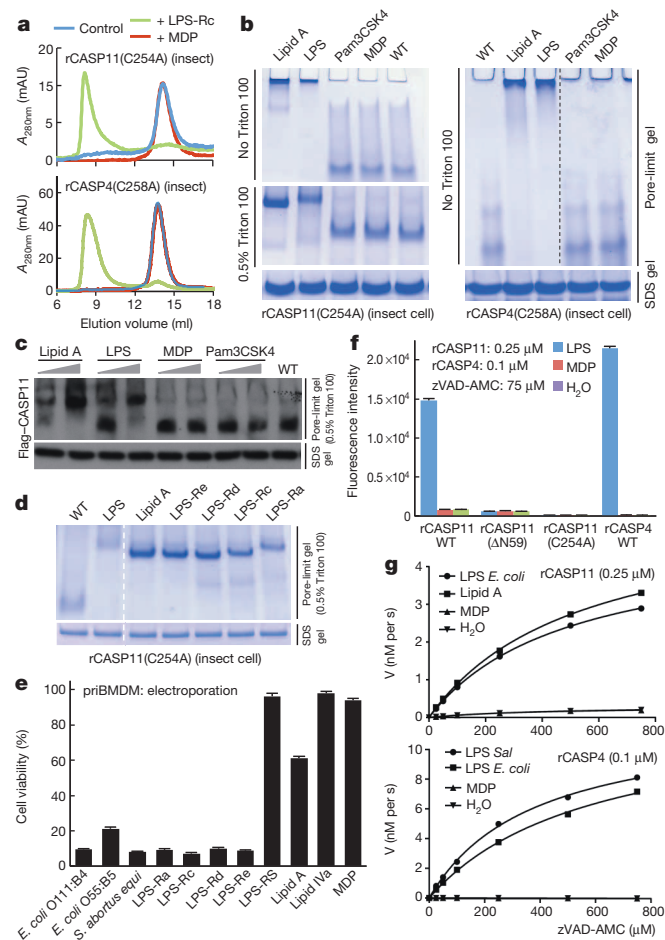


Figure 4 | LPS binding induces oligomerization and activation of caspase-4/11. **a–d**, LPS stimulation of caspase-4/11 oligomerization. Insect-cell purified caspase-11(C254A) and caspase-4(C258A) in **a**, **b**, **d** and Flag tagged caspase-11(C254A) in lysates of transfected 293T cells in **c** were incubated with LPS, LPS variants or other indicated agonists. The samples were analysed by the Superdex 200 gel filtration chromatography in **a** or the pore-limit native gel electrophoresis in **b–d**. **e**, ATP-based cell viability of priBMDMs upon electroporation of LPS or indicated variants. **f**, **g**, LPS binding-induced activation of insect-cell purified caspase-4/11. The caspase activity was determined by measuring the fluorescence intensity of free AMC hydrolyzed from zVAD-AMC. Caspase-11(ΔN59), deletion of the N-terminal 59 residues. The kinetic profile (the initial velocity versus the substrate concentration) is shown in **g**. Data in **e** and **f** are mean values \pm s.d. from three replicates. All data shown are representative of at least three independent experiments.

of LPS into the cell lysates (Extended Data Fig. 6d, e). Thus, the CARD domain in caspase-11 mediates LPS recognition and oligomerization.

LPS stimulates caspase-4/11 activation

Insect-cell purified caspase-4/11 showed a weak basal activity on a generic substrate zVAD-AMC (Fig. 4f, g). Incubation with either LPS or lipid A, but not MDP or Pam3CSK4, stimulated the activity of caspase-11 and caspase-4 approximately 20-fold and several-hundred fold, respectively (Fig. 4f, g and Extended Data Fig. 7a, b). Caspase-11(ΔN59) behaved similarly to the caspase-11 catalytically deficient C254A mutant and showed no activation by LPS (Fig. 4f). Consistent with the oligomerization data (Fig. 4d), LPS-Ra, -Rc, -Rd and -Re all stimulated caspase-4/11 activation (Extended Data Fig. 7c, d). LPS-incubated caspase-4/11 exhibited standard enzyme kinetics with a modest catalytic turnover rate (k_{cat}) (Fig. 4g and Extended Data Fig. 7a, b).

LPS-induced caspase-4 oligomerization was sensitive to Tween-20 detergent (Extended Data Fig. 8a). Increasing Tween-20 concentration from 0.005% to 1% gradually decreased caspase-4 oligomerization, correlating with loss of the catalytic activity (Extended Data Fig. 8a, b).

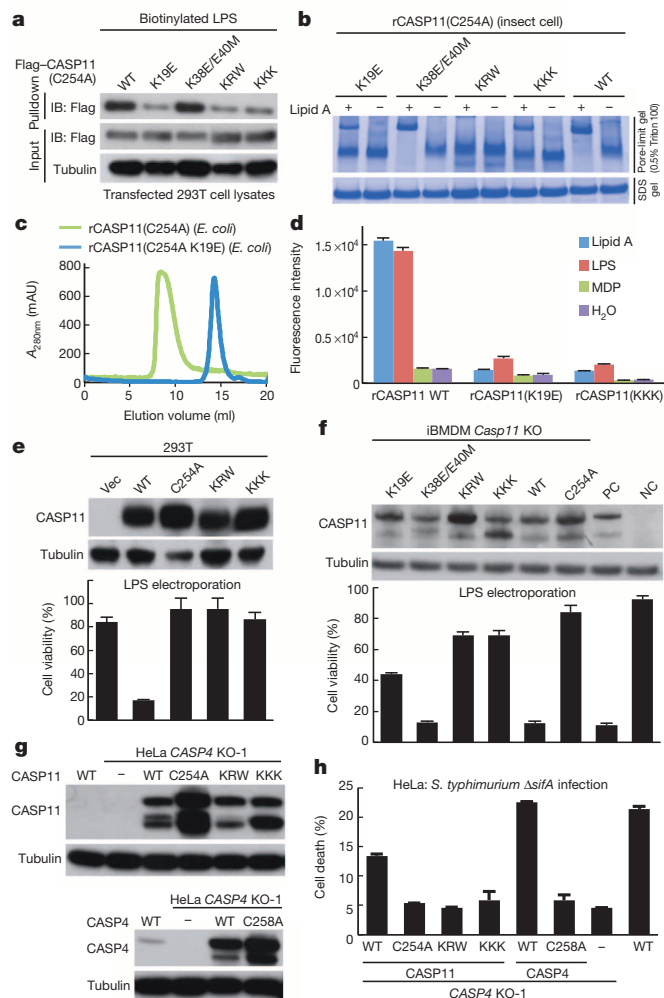


Figure 5 | LPS binding is required for caspase-4/11 oligomerization, activation and induction of pyroptosis. **a**, Streptavidin pull-down of biotin-LPS binding to caspase-11 mutants in transfected 293T cell lysates. **b**, Pore-limit native gel analysis of caspase-11 mutants (from insect cells) oligomerization in response to lipid A. **c**, Gel filtration chromatography of *E. coli*-purified caspase-11 and caspase-11(K19E) (both in a C254A background). **d**, The activity of insect-cell purified caspase-11 mutants following incubation with lipid A, LPS or MDP. **e**, **f**, Assays of caspase-11 mutants mediating LPS electroporation-induced cell death. Wild-type or indicated caspase-11 mutants were stably expressed in 293T cells in **e** and *Casp11*^{-/-} iBMDMs in **f**. Top, anti-caspase-11 immunoblots of total lysates. PC and NC in **f** are WT and *Casp11*^{-/-} iBMDMs, respectively. **g**, **h**, Assays of caspase-11 mutants in mediating *S. typhimurium* Δ*sifA*-induced cell death. *CASP4*^{-/-} HeLa cells expressing wild-type or the indicated caspase-4/11 mutants were infected as indicated. The immunoblots in **g** show the expression of caspase-4/11 mutants. ATP-based cell viability was measured in **e**, **f**, and lactate dehydrogenase release-based cell death was measured in **h**. Data in **d** and **e**, **f**, **h** are mean values \pm s.d. from three replicates. All data shown are representative of at least three independent experiments.

The requirement of oligomerization for catalytic activation was further supported by the properties of lipid IVa and the LPS-RS variant. The penta-acylated LPS-RS from the photosynthetic *R. sphaeroides* is a potent antagonist for LPS-induced toxic shock in mice²¹. Lipid IVa can not activate the non-canonical inflammasome in mice^{6,7}. Notably, biotin–lipid IVa precipitated both cellular and recombinant caspase-4/11 (Fig. 2c, d, h); LPS-RS bound to caspase-11 with a similar affinity to LPS (Figs 2e and 3c, d). However, both analogues failed to stimulate caspase-4/11 oligomerization (Extended Data Fig. 8c) and catalytic activation (Extended Data Fig. 7c, d), consistent with their inability to induce macrophage pyroptosis (Extended Data Fig. 1a and Fig. 4e). Thus, LPS binding-induced oligomerization is a prerequisite for caspase-4/11 activation.

Caspase-11 function requires LPS binding

Known LPS-binding proteins, including *E. coli* FhuA²², *Limulus* haemocyte proteins ALF and Factor C²³, MD2^{24,25} and soluble MD-1²⁶, require basic residues for interacting with the phosphate head groups of lipid A. Through extensive mutagenesis analyses in transfected 293T cells, we identified three caspase-11 mutations, K19E, K52E/R53E/W54A and K62E/K63E/K64E (the latter two are referred to as KRW and KKK, respectively). These residues are all located in the CARD domain and most of them are conserved in caspase-4 but not caspase-11 (Extended Data Fig. 9). These mutations severely disrupted LPS precipitation of full-length caspase-11 and caspase-11 CARD alone (Fig. 5a and Extended Data Fig. 10a), but did not affect overexpression-induced caspase-11 autoproteolysis (Extended Data Fig. 10b). The defective LPS binding was confirmed in the SPR assay (Extended Data Fig. 10c). The caspase-11 mutants, in both a catalytic-active (Cys 254) and inactive (Ala 254) background, were purified from insect cells; the K19E and KKK mutants were severely attenuated and the KRW mutant was completely defective in lipid A induced oligomerization (Fig. 5b). A control K38E/E40M mutant with intact LPS binding (Fig. 5a) responded normally to lipid A induced oligomerization. The three binding-deficient mutants expressed in *E. coli* all eluted as monomers from the gel filtration column (Fig. 5c and data not shown). Insect-cell purified caspase-11(K19E) and KKK mutants were resistant to LPS/lipid A-induced activation (Fig. 5d). The KRW and KKK mutants failed to mediate cytoplasmic LPS-induced death in stable-expression 293T cells (Fig. 5e) and could not restore LPS-induced pyroptosis in *Casp11*^{-/-} iBMDMs (Fig. 5f). The K19E mutant also showed a severe defect. In contrast, the binding-competent K38E/E40M mutant fully rescued LPS-induced iBMDM pyroptosis. The KRW and KKK mutants, in contrast to wild-type caspase-11, could not complement *CASP4*^{-/-} HeLa cells in *S. typhimurium* Δ *sifA*-induced cytotoxicity (Fig. 5g, h). These analyses further support caspase-4/11 directly binding to LPS for activation and induction of necrotic cell death.

Discussion

Inflammatory caspases (caspase-1/4/5/11/12) structurally resemble apoptotic initiator caspases like caspase-8/9. Activation of caspase-1 by death-domain proteins in the inflammasome complex is similar to caspase-9 activation by Apaf-1 in the apoptosome^{27–29}. Caspase-1 and the initiator caspase are activated by signal-induced oligomerization. We show that caspase-4/5/11 directly recognize cytoplasmic LPS, leading to their oligomerization and activation. The mechanism of function of caspase-4/5/11 represents a new paradigm in innate immunity and also a previously unknown mode of caspase activation. The finding that LPS activation of caspase-4 is not limited to human myeloid cells suggests that a re-examination is needed of the mouse model of septic shock. Lastly, factor C and factor G, two serine proteases from horseshoe crab, are also activated by direct binding to LPS and β -(1,3)-D-glucan, respectively. This triggers a coagulation cascade in haemolymphs, serving as an important defence in controlling bacterial infections³⁰. Thus, caspase-4/5/11 are somewhat similar to factor C/G and this group of pattern recognition receptor proteases highlights an evolutionary conserved innate immunity between Merostomata and Mammalia.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 May; accepted 16 July 2014.

Published online 6 August 2014.

1. Kayagaki, N. *et al.* Non-canonical inflammasome activation targets caspase-11. *Nature* **479**, 117–121 (2011).
2. Broz, P. *et al.* Caspase-11 increases susceptibility to *Salmonella* infection in the absence of caspase-1. *Nature* **490**, 288–291 (2012).
3. Akhter, A. *et al.* Caspase-11 promotes the fusion of phagosomes harboring pathogenic bacteria with lysosomes by modulating actin polymerization. *Immunity* **37**, 35–47 (2012).

4. Case, C. L. *et al.* Caspase-11 stimulates rapid flagellin-independent pyroptosis in response to *Legionella pneumophila*. *Proc. Natl Acad. Sci. USA* **110**, 1851–1856 (2013).
5. Achoui, Y. *et al.* Caspase-11 protects against bacteria that escape the vacuole. *Science* **339**, 975–978 (2013).
6. Hagar, J. A., Powell, D. A., Achoui, Y., Ernst, R. K. & Miao, E. A. Cytoplasmic LPS activates caspase-11: implications in TLR4-independent endotoxic shock. *Science* **341**, 1250–1253 (2013).
7. Kayagaki, N. *et al.* Noncanonical inflammasome activation by intracellular LPS independent of TLR4. *Science* **341**, 1246–1249 (2013).
8. Wang, S. *et al.* Murine caspase-11, an ICE-interacting protease, is essential for the activation of ICE. *Cell* **92**, 501–509 (1998).
9. Gurung, P. *et al.* Toll or interleukin-1 receptor (TIR) domain-containing adaptor inducing interferon- β (TRIF)-mediated caspase-11 protease production integrates Toll-like receptor 4 (TLR4) protein- and Nlrp3 inflammasome-mediated host defense against enteropathogens. *J. Biol. Chem.* **287**, 34474–34483 (2012).
10. Rathinam, V. A. *et al.* TRIF licenses caspase-11-dependent NLRP3 inflammasome activation by Gram-negative bacteria. *Cell* **150**, 606–619 (2012).
11. Pilla, D. M. *et al.* Guanylate binding proteins promote caspase-11-dependent pyroptosis in response to cytoplasmic LPS. *Proc. Natl Acad. Sci. USA* **111**, 6046–6051 (2014).
12. Meunier, E. *et al.* Caspase-11 activation requires lysis of pathogen-containing vacuoles by IFN-induced GTPases. *Nature* **509**, 366–370 (2014).
13. Raetz, C. R., Reynolds, C. M., Trent, M. S. & Bishop, R. E. Lipid A modification systems in Gram-negative bacteria. *Annu. Rev. Biochem.* **76**, 295–329 (2007).
14. Whitfield, C. & Trent, M. S. Biosynthesis and export of bacterial lipopolysaccharides. *Annu. Rev. Biochem.* **83**, 99–128 (2014).
15. Zhao, Y. *et al.* The NLRC4 inflammasome receptors for bacterial flagellin and type III secretion apparatus. *Nature* **477**, 596–600 (2011).
16. Lin, X. Y., Choi, M. S. & Porter, A. G. Expression analysis of the human caspase-1 subfamily reveals specific regulation of the CASP5 gene by lipopolysaccharide and interferon- γ . *J. Biol. Chem.* **275**, 39920–39926 (2000).
17. Raymond, A. A. *et al.* Nine procaspases are expressed in normal human epidermis, but only caspase-14 is fully processed. *Br. J. Dermatol.* **156**, 420–427 (2007).
18. Kobayashi, T. *et al.* The *Shigella* OspC3 effector inhibits caspase-4, antagonizes inflammatory cell death, and promotes epithelial infection. *Cell Host Microbe* **13**, 570–583 (2013).
19. Wang, S. *et al.* Identification and characterization of Ich-3, a member of the interleukin-1 β converting enzyme (ICE)/Ced-3 family and an upstream regulator of ICE. *J. Biol. Chem.* **271**, 20580–20587 (1996).
20. Shin, H. J. *et al.* Kinetics of binding of LPS to recombinant CD14, TLR4, and MD-2 proteins. *Mol. Cells* **24**, 119–124 (2007).
21. Qureshi, N., Jarvis, B. W. & Takayama, K. In *Endotoxin in Health and Disease* (eds Brade H., Opal, S. M., Vogel, S. N. & Morrison, D. C.) 687 (Marcel Dekker, 1999).
22. Ferguson, A. D. *et al.* A conserved structural motif for lipopolysaccharide recognition by procaryotic and eucaryotic proteins. *Structure* **8**, 585–592 (2000).
23. Koshiba, T., Hashii, T. & Kawabata, S. A structural perspective on the interaction between lipopolysaccharide and factor C, a receptor involved in recognition of Gram-negative bacteria. *J. Biol. Chem.* **282**, 3962–3967 (2007).
24. Ohto, U., Fukase, K., Miyake, K. & Satow, Y. Crystal structures of human MD-2 and its complex with antiendotoxic lipid IVa. *Science* **316**, 1632–1634 (2007).
25. Park, B. S. *et al.* The structural basis of lipopolysaccharide recognition by the TLR4–MD-2 complex. *Nature* **458**, 1191–1195 (2009) MedlineCrossRef.
26. Yoon, S. I., Hong, M., Han, G. W. & Wilson, I. A. Crystal structure of soluble MD-1 and its interaction with lipid IVa. *Proc. Natl Acad. Sci. USA* **107**, 10990–10995 (2010).
27. Boatright, K. M. & Salvesen, G. S. Mechanisms of caspase activation. *Curr. Opin. Cell Biol.* **15**, 725–731 (2003).
28. McIlwain, D. R., Berger, T. & Mak, T. W. Caspase functions in cell death and disease. *Cold Spring Harb. Perspect. Biol.* **5**, a008656 (2013).
29. Riedl, S. J. & Shi, Y. Molecular mechanisms of caspase regulation during apoptosis. *Nature Rev. Mol. Cell Biol.* **5**, 897–907 (2004).
30. Iwanaga, S. & Lee, B. L. Recent advances in the innate immunity of invertebrate animals. *J. Biochem. Mol. Biol.* **38**, 128–150 (2005).

Acknowledgements We thank Y. Fujimoto and K. Fukase for providing biotin-conjugated lipid A and lipid IVa and B. Finlay for *S. typhimurium* Δ *sifA* strain. We also thank W. Li at Tsinghua University for assistance in SLS and AU analyses and members of the Shao laboratory for discussions and technical assistance. The research was supported in part by an International Early Career Scientist grant from the Howard Hughes Medical Institute and the Beijing Scholar Program to F.S. This work was also supported by the National Basic Research Program of China 973 Program (2012CB518700), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB08020202) and China National Science Foundation Program for Distinguished Young Scholars (31225002) to F.S.

Author Contributions F.S. conceived the study; J.S. and Y.Z. designed and performed the majority of experiments, assisted by Y.W., W.G., J.D., P.L. and L.H. contributed reagents and analytic tools. J.S., Y.Z. and F.S. analysed the data and wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.S. (shaofeng@nibs.ac.cn).

METHODS

Plasmids, antibodies and reagents. cDNAs for *CASP4* and *Casp11* were amplified from reverse-transcribed cDNA from THP1 and C57/BL6 mice-derived primary BMDM cells, respectively. cDNA for *CASP5* was artificially synthesized by the in-house facility. These genes were cloned into a modified pCS2 vector with an N-terminal 3× Flag epitope tag for transient expression in 293T cells, into the FUIGW or pWPI lentiviral vector with an N-terminal Flag tag for stable expression in 293T and immortalized BMDM (iBMDM) cells, into a modified pET vector for recombinant expression in *E. coli*, and into a modified pFastBac vector (Life technologies) for recombinant expression in insect cells. cDNA and expression plasmid for caspase-1 were described previously³¹. The bacterial expression plasmid for *Legionella* SidC protein was prepared in a similar way as that described for SidM³². Truncation mutants were constructed by a standard PCR cloning strategy. cDNAs for the 18 CARD domain-containing proteins were amplified from the Ultimate ORF Clones (Life technologies) and cloned into the pCS2 vector with indicated epitope tags (Extended Data Fig. 3). Point mutations were generated by the QuickChange Site-Directed Mutagenesis Kit (Stratagene). All plasmids were verified by DNA sequencing.

Antibodies for caspase-1 and the Myc epitope were obtained from Santa Cruz Biotechnology. Other antibodies used in this study include HA (Covance), Flag M2 and tubulin (Sigma-Aldrich), rat monoclonal caspase-11 17D9 (Sigma-Aldrich and Santa Cruz), HMGB1 (Epitomics), caspase-5 (Cell Signaling Technology) and mouse monoclonal caspase-4 4B9 (MBL). Homemade anti-ASC antibody was described previously³¹. Biotinylated LPS, Pam3CSK4 and MDP, and unlabelled ultrapure LPS from *E. coli* O111:B4, LPS-RS, Pam3CSK4 and MDP were purchased from InvivoGen. Synthetic lipid IVa was obtained from Peptide Institute, Inc. Synthetic biotin-lipid A and -lipid IVa³³ was kindly provided by Yukari Fujimoto and Koichi Fukase (Osaka University, Japan). LPS from *E. coli* O55:B5 and *S. abortus equi*, Ra, Rc, Rd and Re rough mutants of LPS, lipid A and AMC were all obtained from Sigma-Aldrich. Streptavidin sepharose was from GE Healthcare Life Sciences. The fluorogenic caspase substrate zVAD-AMC was purchased from BACHEM. BS³ and DSS crosslinkers and the LAL Chromogenic Endotoxin Quantitation Kit were obtained from Thermo Scientific. Cell culture products were from Life technologies and all other chemicals were Sigma-Aldrich products unless noted.

Cell culture and bone marrow macrophages. C57BL/6 mice-derived iBMDM cells, HeLa and HaCaT cells were grown in Dulbecco's modified Eagle's medium (DMEM); U937, THP1 and Jurkat T cells were grown in RPMI 1640 medium; HL60 cells were grown in Iscove's modified Dulbecco's medium and HT-29 cells were grown in McCoy's 5a modified medium. All media were supplemented with 10% (vol/vol) fetal bovine serum (FBS) and 2 mM L-glutamine. All cell lines were obtained from ATCC except for the previously described iBMDM¹⁵ and HaCaT cells (China Center for Type Culture Collection, Wuhan University). Cells were grown at 37°C in a 5% CO₂ incubator. JetPRIME (Polyplus Transfection) was used to transfect plasmid DNA into HeLa by following the manufacturer's instruction. 25 ng ml⁻¹ phorbol-12-myristate-13-acetate (PMA) was added to U937 cells for 36–48 h to differentiate the monocytes into macrophages. Differentiated U937 cells were digested with 2 mM EDTA in PBS and sub-cultured in 24-well plates for bacterial infection experiments. Primary BMDMs were prepared from C57BL/6 mice for assaying the caspase-11 pathway by following a standard procedure as previously described^{15,34}. *Ripk3*^{-/-} mice were previously described³⁵. Animal experiments were conducted following the Ministry of Health national guidelines for housing and care of laboratory animals and performed in accordance with institutional regulations after review and approval by the Institutional Animal Care and Use Committee at National Institute of Biological Sciences.

LPS transfection and measurements of cell viability. To stimulate caspase-4/11 noncanonical inflammasome activation, LPS, lipid A or the control agonists (MDP and Pam3CSK4) were electroporated into indicated cells using the Neon Transfection System (Life technologies) following the manufacturer's instructions. Typically, 1 × 10⁶ cells were transfected with 1 µg of the indicated agonists. Cell viability was measured at 10 h after the stimulation for 293T cells and at 2.5 h for other cells. Cell viability was determined by the CellTiter-Glo Luminescent Cell Viability Assay (Promega).

Streptavidin pulldown assay. To assay LPS binding to Flag tagged caspase-4/11 in cell lysates, 293T cells were transfected with indicated caspase-1/4/11 expression plasmids. Cells were collected and lysed in a buffer containing 50 mM Tris-HCl (pH 7.6), 150 mM NaCl, 2 mM EDTA and 1% Triton X-100 supplemented with a protease inhibitor mixture (Roche Molecular Biochemicals) for 30 min. 1 µg of biotin-conjugated lipid A, LPS, Pam3CSK4 and MDP were first immobilized onto 8 µl of streptavidin sepharose beads. Unconjugated ligands were removed by washing the beads for three times with the lysis buffer. The beads were added to pre-cleared lysates to perform the standard pulldown assay. For the competition assay, 20 µg of unlabelled Pam3CSK4, MDP, lipid A, LPS or LPS-RS were first incubated with the pre-cleared lysates for 60 min at 4°C with constant rotation before adding biotinylated LPS-conjugated streptavidin beads. The beads were washed three times with the lysis buffer, and the precipitates were then eluted in 1× SDS sample buffer followed by immunoblotting analyses. To assay LPS binding to endogenous caspase-11, LPS-primed 2 × 10⁷

iBMDM cells were used. To examine the *in vivo* association between LPS and endogenous caspase-11, biotin-conjugated LPS, lipid A or MDP was electroporated into LPS-primed 4 × 10⁷ iBMDM cells. To examine LPS binding to purified recombinant caspase-4/11, 1 µg of insect-cell purified caspase proteins (the catalytic-cysteine mutants) was subjected to a 300-µl binding reaction.

Oligomerization analyses by native gel and gel filtration. Pore-limit native PAGE gel used to analyse LPS binding and LPS-induced protein oligomerization was performed as previously described³⁶ with minor modifications. 4–20% polyacrylamide gels were prepared and run at the constant voltage of 90 V for 8–12 h with buffer (50 mM Tris-HCl (pH 9.0), 7 mM EDTA and 2 mM boric acid) recirculation between the two reservoirs. The AKTA Purifier system (GE Healthcare) was used to perform all gel filtration experiments. Superdex 200 10/300 GL column (GE Healthcare) was used with a flow rate of 0.5 ml min⁻¹ using the buffer containing 50 mM Tris-HCl (pH 7.6), 150 mM NaCl, 2 mM EDTA and 5 mM 2-mercaptoethanol. The molecular weight of eluted proteins was estimated following calibration of the gel filtration column using standard proteins.

RNAi knockdown and transgene stable expression. For siRNA knockdown, HeLa, HT29 and HaCaT were cultured in 6-well plates at 20% confluency at the time of transfection. siRNA transfection was performed using the INTERFERin reagent (Polyplus Transfection) by following the manufacturer's instructions. 5.5 µl of 20 µM siRNA and 10 µl of INTERFERin reagents were used for each well. For THP1 and U937 cells, siRNAs were electroporated into undifferentiated cells at a final concentration of 50 nM. 60 h after transfection, LPS or indicated bacterial ligands were electroporated into the cells to stimulate cell pyroptosis. The siRNA target sequences are TCTACACTATAGTCCAGACCC (*CASP4*-1), GTCTGGACTATAGTGTA TAG (*CASP4*-2), GTATGCCGACAAGCTTGAATT (*CASP3*) and CGTACGCGGA ATACTTCGA (control), which were adopted from those described previously¹⁸. To achieve stable knockdown in U937 cells, pLKO.1-puro or pLKO.1-GFP plasmids expressing shRNAs targeting human *CASP4* or ASC, together with two packing plasmids (pSPAX2 and pMD2G, both from Addgene), were first transfected into 293T cells. Supernatants containing the shRNA-expressing lentivirus were collected 48 h after transfection and used to infect U937 cells for another 48 h. Positive cells were enriched by puromycin selection (*CASP4*-targeting shRNAs) or sorted by flow cytometry (ASC-targeting shRNA). shRNA targeting sequence for ASC (GCCAGGCCT GCACCTTTATA) were adopted from the published literatures^{15,37}. The three shRNAs for *CASP4* (*CASP4*-1, CAAGAGAAGCAACGTATGGCA; *CASP4*-2, CCGAGGAA TGGAGCTGACTTT; *CASP4*-3, AGACTATGTAAAGAAAGAGCT) were from the Sigma-Aldrich TRC library (TRCN0000003513, TRCN0000003510 and TRCN0000003511). The knockdown efficiency was examined by immunoblotting using the corresponding antibodies. Stable gene expression was carried out using the similar protocol as that described for shRNA expression by replacing the pLKO.1 shRNA vector with the pWPI and FUIGW lentiviral gene expressing vectors.

Reverse-transcription PCR (RT-PCR). For RT-PCR analysis, cDNA was prepared as previously described¹⁵. 0.1 µl of the cDNA was used as the template for a 20-µl RT-PCR reaction performed using ExTaq (TaKaRa) on a Bio-Rad PCR System. The primers used for each gene examined are listed below. *CASP4*: CAGACTCTATGCTG AGAGAAGCAACGTATGGCAGGA (forward) and CACCTCTGCAGGCCTGG ACAATGATGAC (reverse); *CASP5*: TAGACTCTTTGCGAAAGAAATCGCGTG GCTCAT (forward) and CACCTCTGCAGGCCTGGACAATGATGAC (reverse); actin: CATGTACGTTGCTATCCAGGC (forward) and CTCCTTAATGTCACGC ACGAT (reverse).

Bacterial lysates preparation. 1 ml of overnight cultures of the indicated bacteria were pelleted, resuspended in 500 µl of PBS and boiled for 10 min. Insoluble components were pelleted by centrifugation at 15,000 r.p.m. for 15 min. The supernatants were collected and used for incubation of insect-cell purified caspase-11 (C254A) protein followed by pore-limit gel electrophoresis analysis.

Purification of recombinant proteins. The insect cell expression of caspase-4 and caspase-11 was achieved with the Bac-to-Bac Baculovirus Expression System (Life technologies) by following the manufacturer's instruction. Briefly, Sf21 insect cells were grown in Sf-900 II SFM at a density of 5 × 10⁵ to 2 × 10⁶ cells per ml. CellFECTIN II Reagent was used to transfect the bacmids harbouring caspase-4/11 into Sf21 cells for recombinant baculovirus production. One litre of cells (1.6 × 10⁶ cells per ml) were infected with 10 ml of the P3 baculovirus and cultured at 28°C for 72 h for large-scale expression. For protein expression in bacteria, *E. coli* BL21 (DE3) strains harbouring the desired expression plasmids were grown in Luria-Bertani medium supplemented with appropriate antibiotics. Protein expression was induced overnight at 22°C with 0.4 mM isopropyl-β-D-thiogalactopyranoside (IPTG) after *D*_{600 nm} (OD_{600 nm}) reached 0.8. To purify the recombinant proteins, insect cells and bacteria were harvested and lysed in a buffer containing 50 mM Tris-HCl (pH 7.6), 300 mM NaCl, 50 mM imidazole and 5 mM 2-mercaptoethanol. For insect cells, 1% (vol/vol) Triton X-100 was added in the lysis buffer. His-tagged proteins were purified by affinity chromatography using Ni-NTA beads (Qiagen). Proteins were eluted with 250 mM imidazole in 50 mM Tris-HCl (pH 7.6) and 300 mM NaCl. Eluted samples were further dialysed

against a buffer containing 50 mM Tris-HCl (pH 7.6) and 150 mM NaCl to remove the imidazole. When needed, proteins were further purified by gel filtration chromatography (GE Healthcare) to achieve homogeneity of the recombinant protein.

Caspase activity and kinetics analysis. Hydrolysis of fluorogenic substrate zVAD-AMC by recombinant caspase proteins was performed in a buffer containing 50 mM HEPES (pH 7.5), 150 mM NaCl, 3 mM EDTA and 0.005% (v/v) Tween-20 and 10 mM DTT. To measure ligand-induced caspase activation, each ligand was incubated with 0.25 μ M caspase-11 or 0.1 μ M caspase-4 proteins (WT or the mutants) in a 100- μ L reaction at 37°C for 30 min. After incubation, zVAD-AMC was added into the reaction at a final concentration of 75 μ M. The reaction mixture was transferred to a 96-well plate and incubated at 37°C for another 30 min. Substrate cleavage was monitored by measuring the emission at 450 nm on excitation at 365 nm on a fluorescent multi-well reader (PerkinElmer EnSpire Multimode Plate Reader). To determine the Michaelis-Menten kinetic parameters, varying substrate concentrations (25–750 μ M) were used for different ligand-incubated caspase-4/11 proteins. After adding the substrate, the reaction was continuously monitored at 20-s intervals for 30 min at 37°C. Only the linear portion of the reaction progress curve was used to determine the initial velocity. The concentration of the cleavage product (free AMC) was determined by using a formulated standard curve, which was generated by plotting the fluorescent intensity versus the concentration of a serially-diluted standard AMC (16–4,000 nM). Kinetic constants were computed by using the GraphPad Prism software (version 5.00).

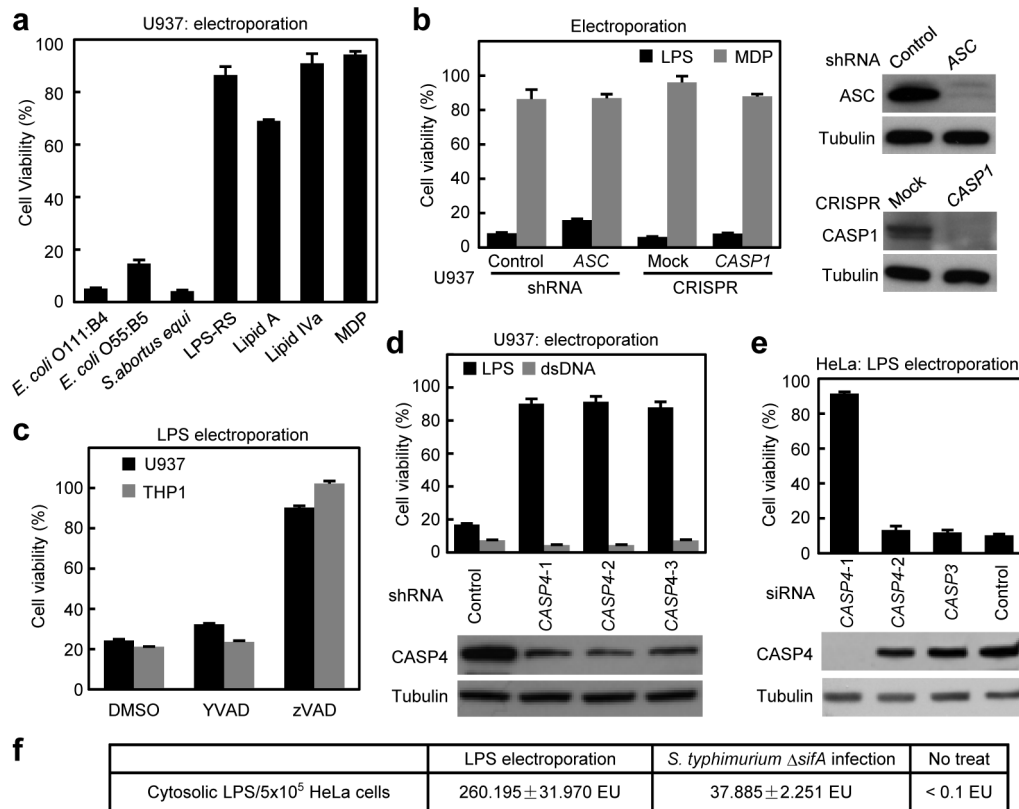
Surface plasmon resonance (SPR). Analyses of ligand binding and binding kinetics were performed at 25°C on a BIAcore T100 SPR instrument (GE Healthcare). The running buffer containing 50 mM HEPES (pH 7.5), 150 mM NaCl, 3 mM EDTA and 0.005% (v/v) Tween-20 was prepared, filtered and degassed before use. Flow cells of CM5 sensor chip were activated for 7 min with a 1:1 mixture of 0.1 M N-hydroxysuccinimide and 0.1 M N-ethyl-N'-(3-diethylaminopropyl)-carbodiimide at a flow rate of 10 μ L min⁻¹. Insect-cell purified caspase-4(C258A) and caspase-11 (C254A) were diluted in 10 mM sodium acetate (pH 5.0) to a concentration of 10 μ g ml⁻¹ and immobilized on different flow cells to ~4,700 (for caspase-4), 4,400 (for caspase-11), 4,200 (for caspase-11(AN59)), 4,500 (for caspase-11(K19E)) and 4,400 (for caspase-11 K38E/E40M, KRW and KKK mutants) response units. Rabbit IgG protein (10 μ g ml⁻¹) was immobilized in 10 mM sodium acetate (pH 4.5) to 4,400 response units. The remaining binding sites on the chips were blocked by 1 M ethanolamine (pH 8.5) at a flow rate of 10 μ L min⁻¹ for 7 min. The ligands were injected at different indicated concentrations and passed over adjacent target and control flow cells at a flow rate of 30 μ L min⁻¹ for 2 min. After 6-min dissociation, the bound analytes were removed by a 20-s wash with 20 mM NaOH. The resulting data after subtracting the control values were analysed by fitting to a 1:1 Langmuir binding model using the BIAcore T100 evaluation software.

Static light scattering and analytic ultracentrifugation. SLS measurements were carried out on a DAWN HELEOS II instrument (Wyatt Technology) at 16°C. Caspase-4/11 proteins were diluted to 2 mg/ml in a buffer containing 50 mM Tris (pH 7.5), 150 mM NaCl, 2 mM EDTA and 2 mM DTT for analyses. Validation of the light scattering detector was performed by using the bovine serum albumin standard. The data were analysed with ASTRA software (Wyatt Technology). Sedimentation velocity analysis was carried out with an XL-I analytical ultracentrifuge (Beckman Coulter) with a four-cell An-60 Ti rotor at 4°C. Caspase-4/11 proteins were prepared in a buffer containing 50 mM Tris (pH 7.5), 150 mM NaCl, 2 mM EDTA and 2 mM DTT at 1 mg ml⁻¹. Buffer alone was used as the reference solution. Data were collected at a speed of 50,000 r.p.m. Absorbance scans were taken at 280 nm at intervals of 4 min in a radial direction. The molecular weights were calculated by the SEDFIT software.

Generation of knockout cell lines. Human codon-optimized Cas9 (hCas9) and GFP-targeting guide RNA (gRNA) expressing vectors (gRNA_GFP-T1)³⁸ were obtained from Addgene. GFP-targeting 19-bp sequence in the gRNA vector was replaced with the desired target sequence by QuickChange Site-Directed Mutagenesis. The target sequences used are GGTGTTTGGATAACTTGG (for *CASP4* KO-1), GTGAGGA AATCTTTGCCCA (for *CASP4* KO-2), CATTCTTCAGTGTGGACCC (for *Casp11*) and TTATCCGTTCCATGGGTGA (for human *CASP1*). To generate the knockout cell lines, 2 μ g of gRNA-expressing plasmid and 6 μ g of hCas9 plasmid together with 1 μ g of pEGFP-C1 vector were electroporated into 6×10^6 iBMDM or U937 cells or transfected into 6×10^6 HeLa cells. 2 or 3 days later, GFP-positive cells were sorted by flow cytometry on the BD Biosciences FACSaria II or Beckman Coulter MoFlo XDP cell sorter. Single clones of sorted cells were obtained by serial dilutions. Cell clones with the desired gene knockout were screened by the T7 endonuclease I assay, checked by sequencing of the PCR fragments, and confirmed by western blot using the corresponding antibodies.

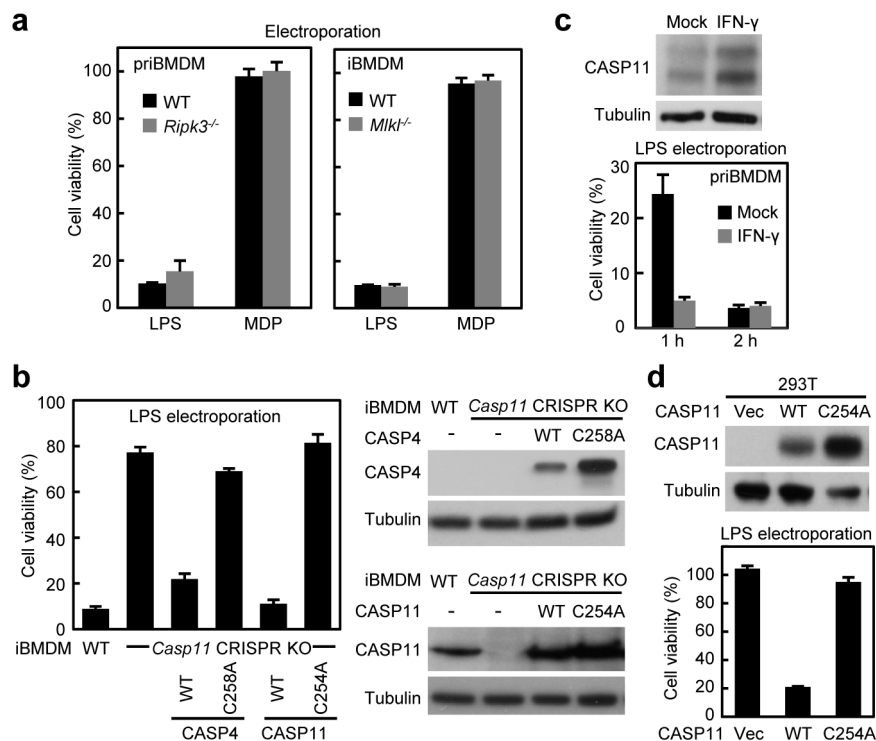
Bacterial strains and infection. *L. pneumophila* strains were cultured on buffered charcoal yeast extract agar plates supplemented with 0.1 mg ml⁻¹ thymidine (CYET). *L. pneumophila* Δ flaA, Δ sdhA and Δ flaA/ Δ sdhA strains were previously described³⁹. For macrophage infection, glycerol stocks of *L. pneumophila* were first streaked on CYET plates and incubated at 37°C in a 5% CO₂ incubator for ~3 to 4 days. Single colonies were then picked and streaked onto new CYET plates. The lawn formed 2 days later was scraped off, diluted in sterile water, and added to cells at a multiplicity of infection (MOI) of 20. Infection was facilitated by a centrifugation of 1000g for 10 min. For *S. typhimurium* infection, overnight culture of wild-type *S. typhimurium* str. SL1344 or Δ sifA mutant strain⁴⁰ (kindly provided by Brett Finlay, University of British Columbia, Canada) were added to cells at a MOI of 20, centrifuged for 10 min at 800g, and then incubated at 37°C for another 1.5 h. Extracellular bacterial growth was stopped by adding 60 μ g ml⁻¹ gentamicin. Infection-induced cell death was measured 4 h later for *L. pneumophila* and 6 h later for *S. typhimurium* by the lactate dehydrogenase (LDH) assay using CytoTox 96 Non-Radioactive Cytotoxicity Assay kit (Promega). All infection assays were performed in the media without serum and antibiotics.

- Gong, Y. N. *et al.* Chemical probing reveals insights into the signaling mechanism of inflammasome activation. *Cell Res.* **20**, 1289–1305 (2010).
- Zhu, Y. *et al.* Structural mechanism of host Rab1 activation by the bifunctional Legionella type IV effector SidM/DrrA. *Proc. Natl Acad. Sci. USA* **107**, 4699–4704 (2010).
- Fujimoto, Y. *et al.* Synthesis of lipid A and its analogues for investigation of the structural basis for their bioactivity. *J. Endotoxin Res.* **11**, 341–347 (2005).
- Yang, J., Zhao, Y., Shi, J. & Shao, F. Human NAIp and mouse NAIp1 recognize bacterial type III secretion needle protein for inflammasome activation. *Proc. Natl Acad. Sci. USA* **110**, 14408–14413 (2013).
- He, S. *et al.* Receptor interacting protein kinase-3 determines cellular necrotic response to TNF- α . *Cell* **137**, 1100–1111 (2009).
- Boatright, K. M. *et al.* A unified model for apical caspase activation. *Mol. Cell* **11**, 529–541 (2003).
- Li, H., Willingham, S. B., Ting, J. P. & Re, F. Cutting edge: inflammasome activation by alum and alum's adjuvant effect are mediated by NLRP3. *J. Immunol.* **181**, 17–21 (2008).
- Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
- Ge, J., Gong, Y. N., Xu, Y. & Shao, F. Preventing bacterial DNA release and absent in melanoma 2 inflammasome activation by a Legionella effector functioning in membrane trafficking. *Proc. Natl Acad. Sci. USA* **109**, 6193–6198 (2012).
- Brumell, J. H., Tang, P., Zaharik, M. L. & Finlay, B. B. Disruption of the Salmonella-containing vacuole leads to increased replication of *Salmonella enterica* serovar typhimurium in the cytosol of epithelial cells. *Infect. Immun.* **70**, 3264–3270 (2002).



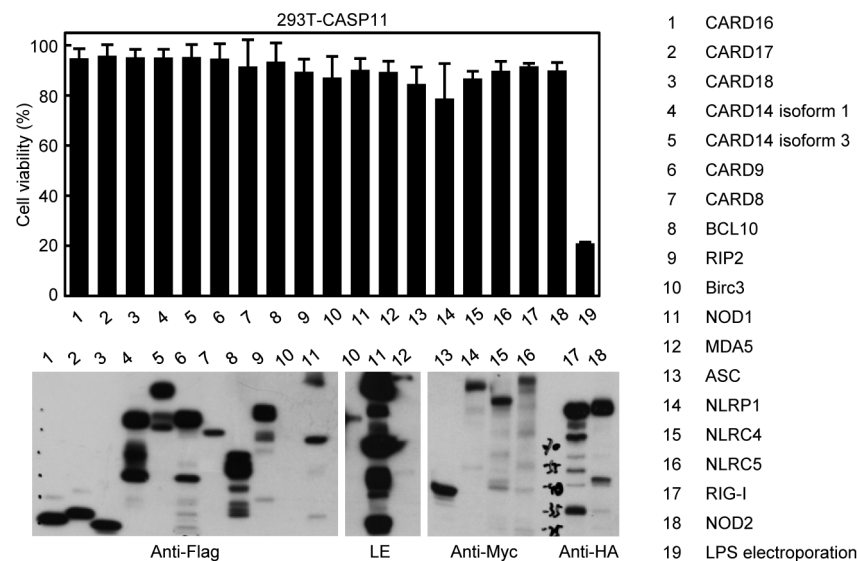
Extended Data Figure 1 | Cytoplasmic LPS induces ASC/caspase-1-independent but caspase-4-dependent necrosis in human cells. **a**, Viability of U937 cells upon electroporation with lipid A, MDP, LPS-RS or wild-type LPS derived from the indicated bacterial strains. **b**, Effects of ASC knockdown and *CASP1* knockout on LPS electroporation-induced pyroptosis in U937 cells. ASC knockdown U937 cells were generated by lentivirus-mediated shRNA transduction. *CASP1* knockout U937 cells were generated by CRISPR/Cas9-mediated targeting. **c**, Effects of caspase-1 (YVAD) or pan-caspase inhibitor (zVAD) on LPS electroporation-induced cytotoxicity in U937 and THP1 cells. **d**, **e**, Effects of *CASP4* knockdown on LPS electroporation-induced cytotoxicity in U937 and HeLa cells. Control or a *CASP4*-specific shRNA (*CASP4*-1/2/3)

was stably expressed in U937 cells in **d**; a *CASP4*-specific (*CASP4*-1/2) or *CASP3*-targeting or control siRNA was transfected into HeLa cells in **e**. **f**, Quantification of cytosolic LPS upon electroporation or bacterial infection. 5×10^5 *CASP4*^{-/-} HeLa cells were either electroporated with 0.5 μ g of LPS or infected with *S. typhimurium* Δ *sifA*. Cells were washed with cold PBS 5 times and lysed in a buffer containing 1% Triton X-100. ATP-based cell viability was measured in **a–e**. The immunoblots show the knockdown of ASC and *CASP4* in **b**, **d**, **e** and knockout of *CASP1* in **b**. Graphs show the mean values \pm s.d. from three technical replicates. All data shown are representative of at least three independent experiments.



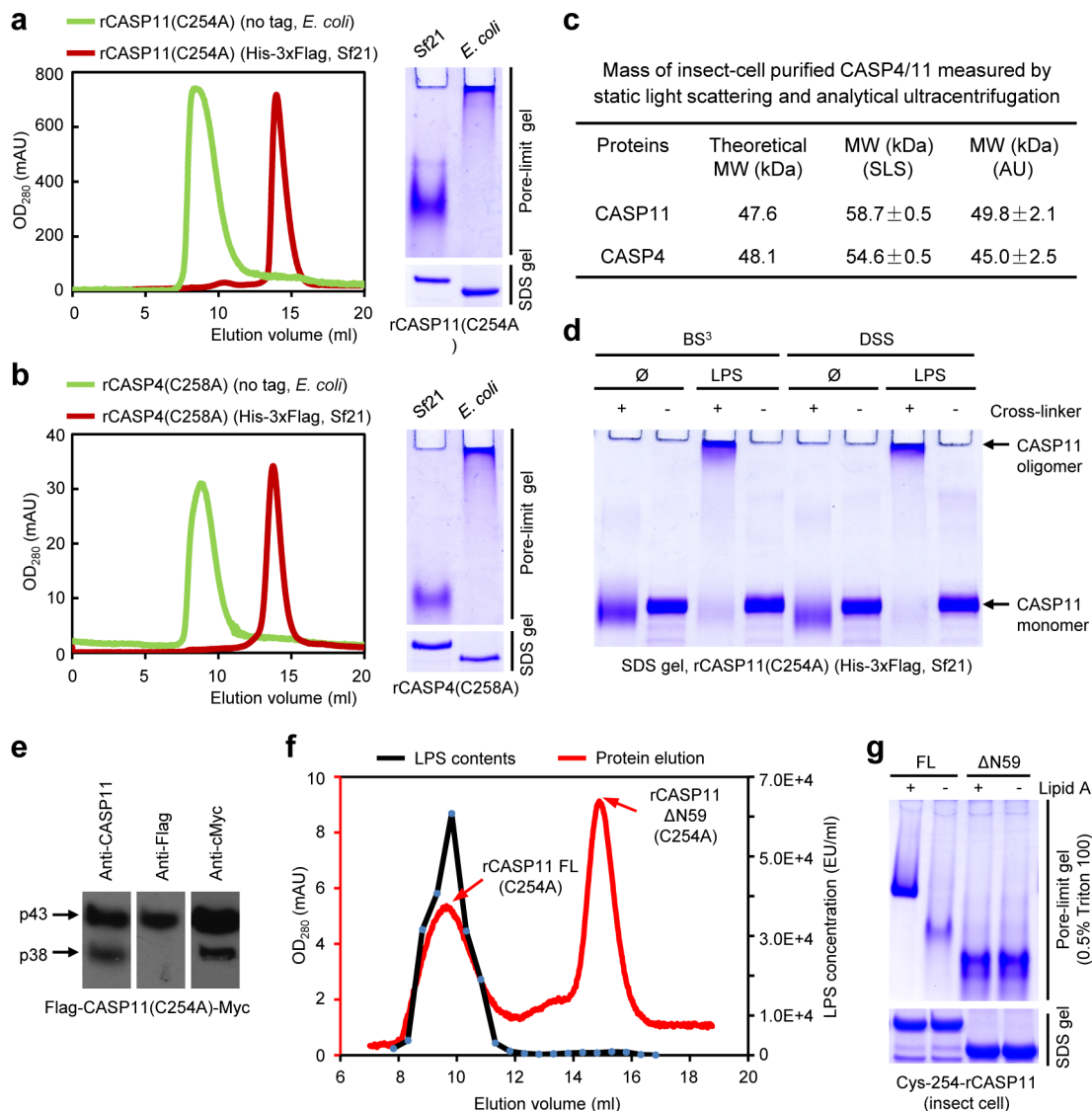
Extended Data Figure 2 | Caspase-11-mediated cell death in response to cytoplasmic LPS and complementation by caspase-4. **a**, Effects of *Ripk3* and *Mlkl* deficiency on LPS-induced necrosis in mouse BMDMs. **b**, Wild-type (WT) or the catalytically inactive caspase-11(C254A) or caspase-4(C258A) mutant was stably expressed in *Casp11*^{-/-} iBMDM cells generated by CRISPR/Cas9-mediated targeting. **c**, IFN-γ stimulates caspase-11 expression and increase the sensitivity of primary BMDMs to LPS electroporation.

d, Wild-type (WT) or the catalytically inactive caspase-11(C254A) mutant was stably expressed in 293T cells. LPS was delivered by electroporation. ATP-based cell viability in **a–d** was expressed as mean values ± s.d. from three technical replicates. The accompanying immunoblots show the expression of caspase-4 or caspase-11. Data shown in **a**, **c** are representative of two independent experiments and those in other panels are of at least three independent experiments.



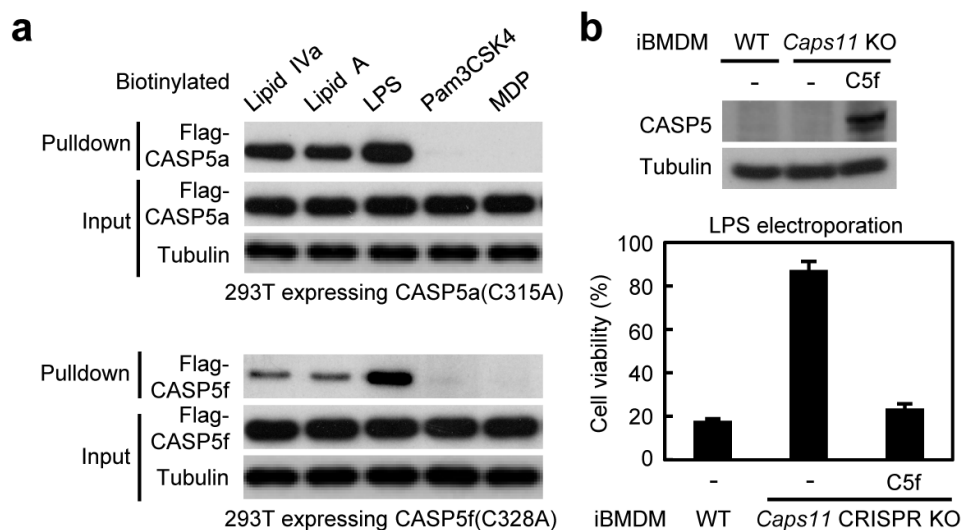
Extended Data Figure 3 | Assays of the ability of known CARD-domain proteins in inducing caspase-11 activation. Each of the 18 indicated CARD-domain proteins was overexpressed in 293T cells stably expressing caspase-11. As a control (#19), 293T-caspase-11 cells were subjected to LPS electroporation. The top shows the ATP-based cell viability expressed as mean

values \pm s.d. from three technical replicates. Anti-Flag, Myc and HA immunoblots confirm the expression of indicated CARD-domain proteins. LE, long exposure. Data shown are representative of three independent experiments.



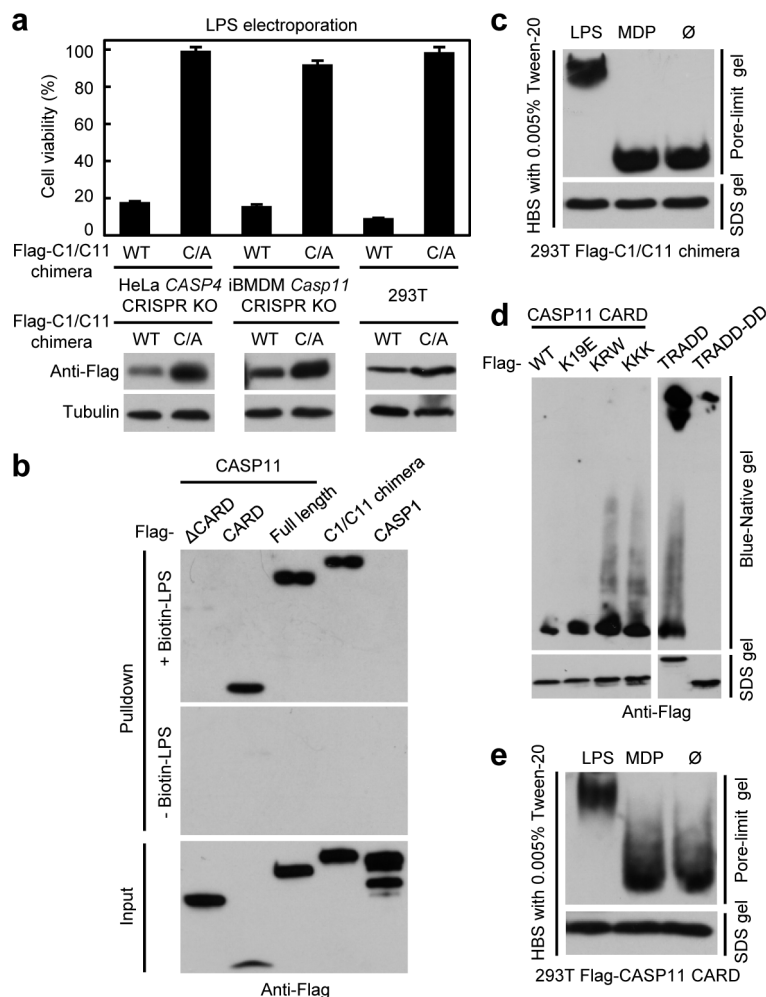
Extended Data Figure 4 | LPS stimulates the oligomerization of caspase-4/11 in the CARD domain-dependent manner. **a, b**, Recombinant caspase-4/11 purified from *E. coli* and insect cells exhibit different oligomerization states. The left and right show the gel filtration chromatography and Coomassie blue stained pore-limit native gels of indicated recombinant caspase-11(C254A) (**a**) and caspase-4(C258A) protein (**b**), respectively. **c**, Molecular weight determination of insect-cell purified caspase-4 and caspase-11 by static light scattering (SLS) and analytic ultracentrifugation (AU). Shown are mean values ± s.d. **d**, Responses of insect-cell purified caspase-11 (before and after LPS incubation) to crosslinking agents (BS³ and DSS). Shown are Coomassie blue-stained SDS-PAGE gels. **e**, Immunoblotting of Flag tagged caspase-11-Myc expressed in 293T cells. **f**, Gel filtration chromatography of LPS-incubated

full-length (FL) and the N-terminal 59-residue deletion (ΔN59) mutant caspase-11. Insect-cell purified caspase-11 (FL and ΔN59, both in the C254A background) was incubated with LPS-Rc overnight at 4°C. The proteins were further purified by affinity chromatography using Ni-NTA beads and subjected to Superdex 200 gel filtration chromatography. The LPS contents in each elution fraction (0.5 ml) were plotted against the corresponding elution volume. **g**, Pore-limit native gel analysis of lipid A induced oligomerization of insect-cell purified caspase-11 FL and the ΔN59 mutant (both in the catalytic-cysteine intact background). Data shown in **c, f**, are representative of two independent experiments and those in other panels are of at least three independent experiments.



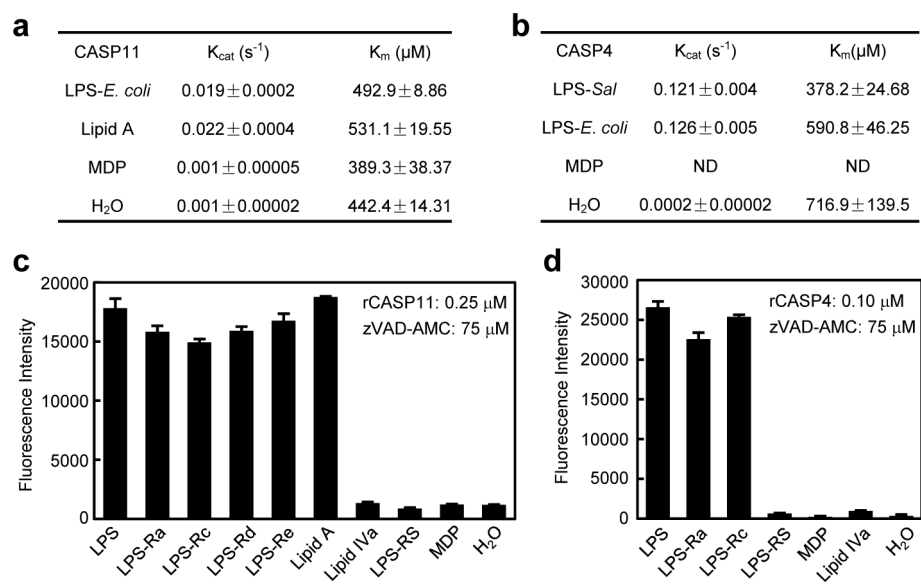
Extended Data Figure 5 | Caspase-5 can bind to LPS and complement *Casp11*^{-/-} iBMDMs in cytoplasmic LPS-induced cell death. **a**, Streptavidin pulldown assays of biotin-conjugated lipid A, lipid IVa, LPS, Pam3CSK4, and MDP binding to Flag tagged CASP5a (C315A) and CASP5f (C328A) in transfected 293T cell lysates. **b**, Caspase-5 complementation of *Casp11*^{-/-}

iBMDMs in cytoplasmic LPS-induced pyroptosis. Wild-type caspase-5f (C5f) was stably expressed in *Casp11*^{-/-} iBMDMs generated by CRISPR/Cas9-mediated targeting. ATP-based cell viability was expressed as mean values \pm s.d. from three technical replicates. All data shown are representative of at least three independent experiments.



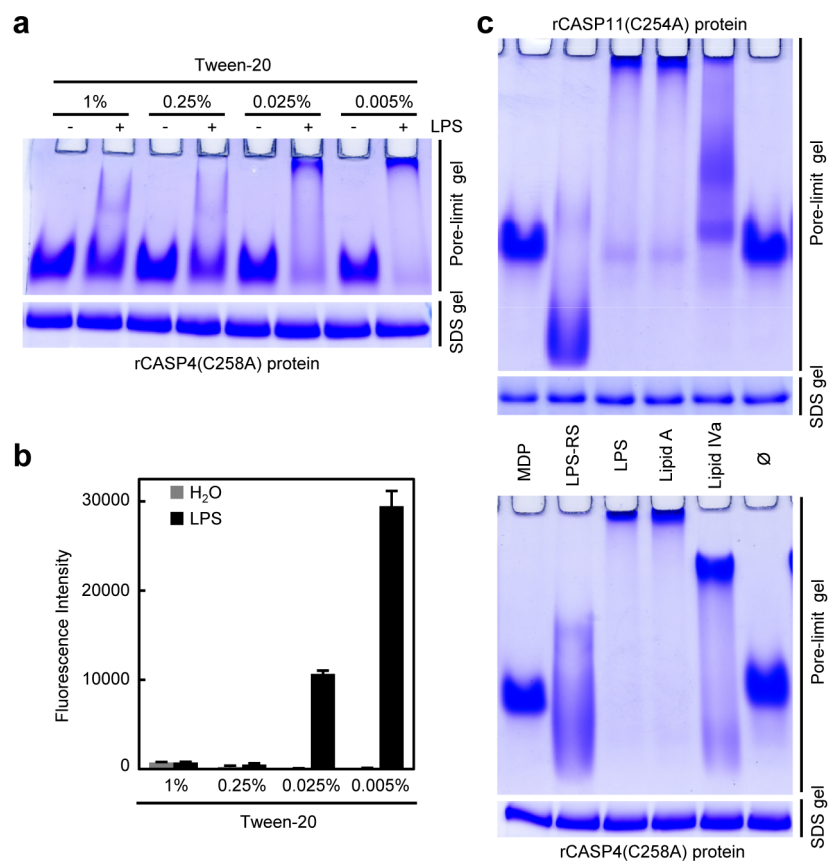
Extended Data Figure 6 | LPS binds to caspase-11 CARD domain and induces its oligomerization. **a**, Assays of the caspase-1/11 CARD-domain chimaeric protein in sensing cytoplasmic LPS. The chimaeric protein (Flag tagged C1/C11) was generated by replacing the CARD domain in Flag tagged caspase-1 with that of caspase-11. *CASP4* knockout HeLa cells, *Casp11* knockout iBMDM cells and 293T cells stably expressing Flag-C1/C11 (WT or the catalytic cysteine mutant (C/A)) were electroporated with LPS. ATP-based cell viability was expressed as mean values \pm s.d. from three technical replicates. Expression of the chimaeric proteins were shown by anti-Flag immunoblotting. **b**, Streptavidin pulldown assays of the binding of biotinylated LPS to the C1/C11 chimaera or indicated caspase-11 variants in lysates of

transfected 293T cells. **c**, **e**, LPS-induced oligomerization of the Flag tagged C1/C11 chimaera and the caspase-11 CARD domain itself. Lysates of 293T cells stably expressing the chimaeric protein (**c**) or caspase-11 CARD domain (**e**) were subjected to incubation with LPS and then analysed on pore-limit native gels. Cells were lysed in the HBS buffer (50 mM HEPES, pH 7.5, 150 mM NaCl and 3 mM EDTA) supplemented with 0.005% Tween-20. **d**, Oligomerization assays of caspase-11 CARD domain overexpressed in 293T cells. TRADD and TRADD death domain (DD) were included as positive controls and Blue Native gel was employed to examine the oligomerization. Shown in **b-e** are anti-Flag immunoblots. All data shown are representative of at least three independent experiments.



Extended Data Figure 7 | LPS binding-induced activation of caspase-4 and caspase-11. Insect-cell purified full-length caspases were subjected to incubation with MDP, lipid A, lipid IVa, LPS or the indicated LPS variants. The caspase activity was determined by measuring the fluorescence intensity of free AMC hydrolyzed from the zVAD-AMC substrate. Shown in **a**, **b** are the

Michaelis–Menten kinetic parameters (k_{cat} and K_m) (values \pm standard errors; ND, not determined). Graphs in **c**, **d**, show fluorescence intensity values determined with indicated enzyme and substrate concentration as mean values \pm s.d. from three replicates. All data shown are representative of at least three independent experiments.



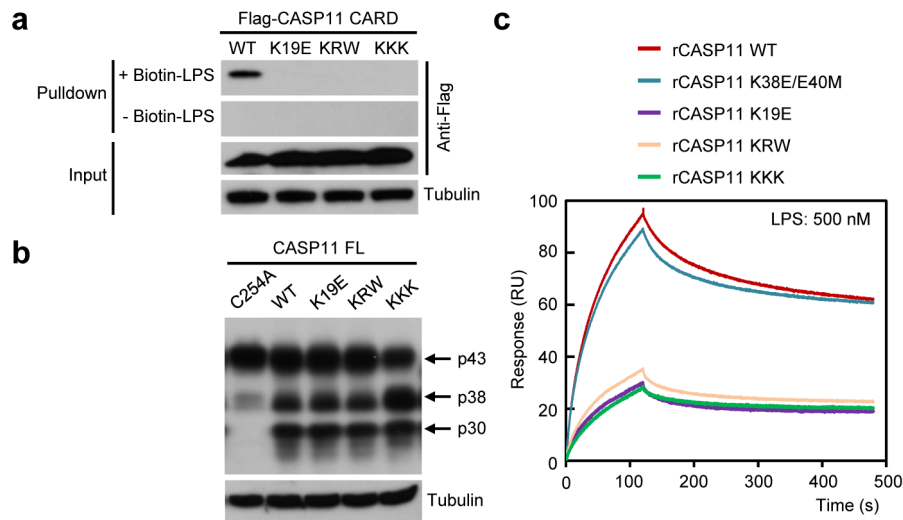
Extended Data Figure 8 | LPS binding-induced oligomerization is required for caspase-4/11 activation. **a, b,** Effects of Tween-20 on LPS binding-induced oligomerization and catalytic activation of caspase-4. LPS-incubated recombinant caspase-4 was subjected to further incubation with increasing concentrations of Tween-20 as indicated. The fluorescence intensities in **b** are mean values \pm s.d. from three replicates. **c,** Assays of lipid IVa and LPS-RS in

stimulating oligomerization of caspase-4/11. Insect-cell purified caspase-11(C254A) and caspase-4(C258A) were subjected to incubation with indicated agonists, and the samples were analysed by the pore-limit native PAGE gel electrophoresis. All data shown are representative of at least three independent experiments.

CASP11_MOUSE	1	MAENKHPDKPLKVLEQLGKEVLTEYLEKLVSQSNVLKLEEDKQKFNNAESDKRWFVFDAMKK
CASP4_HUMAN	1	MAEGNHRKKPLKVLESLEKDFLTGVLDNLVEQNVLNWKEEEKKKYYDAKTBDKVRVMADSMQE
CASP4_CHIMPANZEE	1	MAEGNHRKKPLKMLESLEKDFLTGVLDNLVEQNVLNWKEEEKKKYYDAKTBDKVRVMADSIQE
CASP4_SHEEP	1	MAEDKHKNKPLKTLESLEKELISGLDDFVEKNVLEKEEEKKKIYDAKLQDKARVLIDSIRQ
CASP4_BOVIN	1	MAEDKHKNKPLKMLESLEKELISGLDDFVEKNVLEKEEEKKKIYDAKLQDKARVLVDSIRQ
CASP4_PIG	1	MAEVQKKNPPLKILESMGKELITGVLDLLEKDVLEKEEEKKNIYDAKLQDKARILMDSVLQ
CASP11_MOUSE	64	KHSKVGEMLLQTFFSVDPGSHHGEANLEME...EPPEESLNTLKLCSPEEETRLCREKTQEIY
CASP4_HUMAN	64	KQRMAGQMLLQTFFNIDQISPNKKAHPNMEAGPPESGESTDALKLCPHEEFLRLCKERAEEIY
CASP4_CHIMPANZEE	64	KQRMAGQMLLQTFFNIDQISPNKKAHPNMEAGPPESGESTDALKLCPHEEFLRLCKERAEEIY
CASP4_SHEEP	64	KNQEAQGVFVQTFNLNIDKNSTNIKAPEETVAGPDESAGSAATLKLCPHEEFLRLCKERAGEIY
CASP4_BOVIN	64	KNQEAQGVFVQTFNLNIDKNSTNIKAPEETVAGPDESAGSAATLKLCPHEEFLRLCKERAGEIY
CASP4_PIG	64	KRHEASQVFVKTFNLNMDKNSTNIQAPEKIMAGPDESEESTDTLKLCPHEEFLRLCKERTGEIY
CASP11_MOUSE	123	PIKEANGRTRKALIIICNTEFKHLSLRYGANFDITGMKGLEEDLGVDVVKBELTAEGMESEMK
CASP4_HUMAN	127	PIKERNNRTRLALIIICNTEFDHLPFRNGADFDITGMKELLEGLDYSVDVEENLTARDMESALR
CASP4_CHIMPANZEE	127	PIKERNNRTRLALIIICNTEFDHLPFRNGADFDITGMKELLEGLDYSVDVEENLTARDMESALR
CASP4_SHEEP	127	PIKERKDRTRLALIIICNTEFDHLPFRNGADLDILGMKQLEGLGYTVVEBEKLTARDMESVLW
CASP4_BOVIN	127	PIKERKDRTRLALIIICNTEFDHMPFRNGAALDILGMKQLEGLGYTVVEBEKLTARDMESVLW
CASP4_PIG	127	PIKNKDRTRLALIIICNTEFDHLPFRNGADLDIKGMKQLEGLGYSDVVEBEKLTAKEMESVLW
CASP11_MOUSE	186	DFAALSEHQTSNSTFLVLMSHGTLHGICGTMHSEKTEPDVLQYDTIYQIFNNCHCPGLRDKPKV
CASP4_HUMAN	190	AFATRPEHKSSDSTFLVLMSHGTLHGICGTVHDEKKPDVLLYDTIFQIFNNRNCLSLKDKPKV
CASP4_CHIMPANZEE	190	AFATRPEHKSSDSTFLVLMSHGTLHGICGTVHDEKKPDVLLYDTIFQIFNNRNCLSLKDKPKV
CASP4_SHEEP	190	KFAAREBEHKSSDSTFLVFMHSGTLDGICGTMHSEEBPDVLPYDTIFRTFNNRNCLSLKDKPKV
CASP4_BOVIN	190	KFAAREBEHKSSDSTFLVFMHSGTLDGICGTMHSEEBPDVLPYDTIFRTFNNRNCLSLKDKPKV
CASP4_PIG	190	GFAARQEHKSSDSTFLVFMHSGTLDGICGTMHSDKKPDVLLHYDTIFKIFNNRNCLGLKDKPKV
CASP11_MOUSE	249	IIVQACRGGNSGEMWIRESSKQPQLCRGVLDLPRNMEADAVKLSHVEKDFIAFYSTTPHHLSYRD
CASP4_HUMAN	253	IIVQACRGANRGELWVRDSPASLEVASSQSSENLEEDAVYKTHVEKDFIAFGSSTPHNVSWRD
CASP4_CHIMPANZEE	253	IIVQACRGANRGELWVRDSPASLEVASSQSSENLEEDAVYKTHVEKDFIAFGSSTPHNVSWRD
CASP4_SHEEP	253	IIVQACRGANRGELWVSDSPAALADSFSSQSPENLEEDAVYKTHVEKDFIAFGSSTPHNVSWRD
CASP4_BOVIN	253	IIVQACRGANRGELWVSDSPAALADSFSSQSPENLEEDAVYKTHVEKDFIAFGSSTPHNVSWRD
CASP4_PIG	253	IIVQACRGANRGGEVWVSDSPATLADSSSESPEDLEEDAVYKTHVEKDFIAFGSSTPHNVSWRD
CASP11_MOUSE	312	KTGGSYFITRLISCFRKHACSCHLFDIFLKVQSSFEKASIHSQMPTIDRATLTRYFYLFPGN
CASP4_HUMAN	316	STMGSIFITQLITCFQKYSWCCHLEEVFRKVQSSFETPRAKAQMPTIERLSMTRYFYLFPGN
CASP4_CHIMPANZEE	316	STMGSIFITQLITCFQKYSWCCHLEEVFRKVQSSFETPRAKAQMPTIERLSMTRYFYLFPGN
CASP4_SHEEP	316	IKKGSIFITRLITCFQKYAWCCHLEEVFRKVQSSFEKPNVKAQMPTVERLSMTRYFYLFPGN
CASP4_BOVIN	316	IKKGSIFITRLITCFQKYAWCCHLEEVFRKVQSSFEKPNVKAQMPTVERLSMTRYFYLFPGN
CASP4_PIG	316	VTKGSTFETQLITCFQKYAWRCHLEEVFRKVQSSFETPSVKAQMPTIERLSMTRYFYLFPGN

Extended Data Figure 9 | Multiple sequence alignment of caspase-11 and caspase-4 derived from various mammals. Indicated caspase-sequences were aligned by using the ClustalW2 algorithm and the alignment was generated in

ESPrpt 3.0 (<http://esprpt.ibcp.fr/ESPrpt/cgi-bin/ESPrpt.cgi>). Residues important for LPS binding to caspase-11 are highlighted in orange.



Extended Data Figure 10 | Point mutations in caspase-11 CARD domain disrupt LPS binding to both caspase-11 CARD domain and full-length caspase-11. **a**, Streptavidin pulldown assays of the binding of biotinylated LPS to wild-type or indicated point mutants of Flag tagged caspase-11 CARD domain in lysates of transfected 293T cells. **b**, Effects of the CARD-domain point mutants on transient overexpression-induced caspase-11 autoprocessing

in 293T cells. **c**, Surface plasmon resonance measurements of the binding between LPS and the CARD-domain point mutants of insect-cell purified caspase-11. Colour indicated caspase-11 proteins were immobilized on the chips and shown are the corresponding sensorgrams expressed in RU (response unit) versus time after subtracting the control signal. All data shown are representative of two independent experiments.

Crystal structure of a eukaryotic group II intron lariat

Aaron R. Robart¹, Russell T. Chan¹, Jessica K. Peters¹, Kanagalaghatta R. Rajashankar² & Navtej Toor¹

The formation of branched lariat RNA is an evolutionarily conserved feature of splicing reactions for both group II and spliceosomal introns. The lariat is important for the fidelity of 5' splice-site selection and consists of a 2'-5' phosphodiester bond between a bulged adenosine and the 5' end of the intron. To gain insight into this ubiquitous intramolecular linkage, we determined the crystal structure of a eukaryotic group IIB intron in the lariat form at 3.7 Å. This revealed that two tandem tetraloop-receptor interactions, η - η' and π - π' , place domain VI in the core to position the lariat bond in the post-catalytic state. On the basis of structural and biochemical data, we propose that π - π' is a dynamic interaction that mediates the transition between the two steps of splicing, with η - η' serving an ancillary role. The structure also reveals a four-magnesium-ion cluster involved in both catalysis and positioning of the 5' end. Given the evolutionary relationship between group II and nuclear introns, it is likely that this active site configuration exists in the spliceosome as well.

Splicing of nuclear introns results in the formation of circular RNAs¹ having a branched lariat structure containing an unusual 2'-5' phosphodiester bond^{2,3}. This branched RNA product has also been found in group II introns^{4,5}, which are self-splicing ribozymes. Defects in lariat formation result in aberrant splicing and human disease⁶. In eukaryotes, splicing of nuclear introns is catalysed by a large ribonucleoprotein complex called the spliceosome, which is thought to share a common ancestor with group II introns^{7,8}.

Group II introns are catalytic RNAs with six structural domains (Extended Data Fig. 1) that splice via two transesterification reactions. In the first step of splicing, the 2'-OH of a bulged adenosine residue is activated for nucleophilic attack at the 5' splice site to generate lariat RNA^{4,5}. In the second step, the 3'-OH of the 5' exon attacks the 3' splice site to form ligated exons and excised intron lariat. The highly conserved domain V (DV) forms the group II intron active site by binding catalytic metal ions⁹, and domain VI (DVI) contains the bulged adenosine used as the nucleophile in the first step of splicing¹⁰.

Group II introns are divided into three structural classes: IIA, IIB and IIC^{11,12}. Historically, the two model systems used to study group II intron structure and function have been 'canonical' eukaryotic IIB introns: *P.li.LSUI2* from the brown algae *Pylaiella littoralis*¹³ and *al5 γ* from the yeast *Saccharomyces cerevisiae*⁶. However, the only available crystal structure is of a IIC representative from the bacterium *Oceanobacillus iheyensis*⁹. This idiosyncratic IIC intron class is the most primitive¹⁴ and splices through hydrolysis to form linear intron¹⁵. In contrast, eukaryotic IIA and IIB introns form lariat, are evolutionarily later branching¹⁴, and therefore more closely related to the spliceosome. We targeted the *P.li.LSUI2* intron for structure determination since it contains a functional DVI that forms large amounts of lariat during splicing¹³.

Overall structure

Here we present the structure of the *P.li.LSUI2* intron in the post-catalytic lariat form with ligated exon product at 3.7 Å resolution (Extended Data Table 1a) solved using a Yb³⁺ derivative (Extended Data Fig. 2). This represents the first crystal structure of a 2'-5' branched RNA molecule.

Reflecting the ability of IIB introns to form lariat, there are a multitude of unique tertiary interactions in the *P.li.LSUI2* intron compared

to the *O. iheyensis* structure (Fig. 1). These newly visualized contacts include EBS2-IBS2 (Extended Data Fig. 3), μ - μ' , ϵ - ϵ' and the canonical form of κ - κ' . Unlike the *O. iheyensis* structure, domains II and III interact with multiple domains through long-range interactions to stabilize the overall fold of the *P.li.LSUI2* intron. We can now visualize the location of DVI within the intron structure (Fig. 1 and Extended Data Fig. 4), as well as the 2'-5' lariat linkage between the first residue and the bulged adenosine.

Newly visualized tertiary interactions

One of the most highly conserved tertiary contacts in group II introns is the κ - κ' interaction between the base of the catalytic DV stem and domain I (DI)¹⁶. The conserved κ sequence GAA, nucleotide A171 from near the κ region, and residues from a GUAAAC pentaloop in DIII converge to form a pentuple adenosine base stack (underlined residues) that inserts into the minor groove at the base of DV, rigidly placing the active site into the DI scaffold (Fig. 2a and Extended Data Fig. 5a).

The ϵ - ϵ' interaction is critical for catalysis, with disruption through mutagenesis resulting in complete loss of splicing activity¹⁷. This interaction consists of nucleotides G106 and C107 pairing with C4 and G3 from the 5' end of the intron (Fig. 2b and Extended Data Fig. 5b). The end result of these contacts is the formation of five conserved bases stacking in the following order (from bottom to top): A573, U2, G5, C4 and G3. This anchors the 5' end in the core of the intron.

The conserved GUAA linker connecting domains I and II (J1/2) adopts an unusual backbone configuration that interacts with the 5' end, the junction between domains II and III (J2/3), and DIII (Fig. 2c and Extended Data Fig. 5c). The 5' end and J2/3 directly interact with the active site through J1/2 positioning these regions to stabilize the core. This new long-range contact (designated as ρ - ρ') consists of two adenosines from J1/2 docking into the basal stem of DIII (Fig. 2c).

We can now correlate the function of DIII as a catalytic effector in group II introns. DIII interacts with the intron core through the aforementioned GUAAAC pentaloop, which docks into the base of DV (Fig. 2d). This μ - μ' interaction¹⁸ serves to buttress the opposite side of DV from where catalysis takes place. Furthermore, a GAAA tetraloop from DIII interacts with the base of the DII stem (designated as τ - τ') to provide

¹Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, California 92093, USA. ²NE-CAT and Department of Chemistry and Chemical Biology, Cornell University, Argonne National Laboratory, Argonne, Illinois 60439, USA.

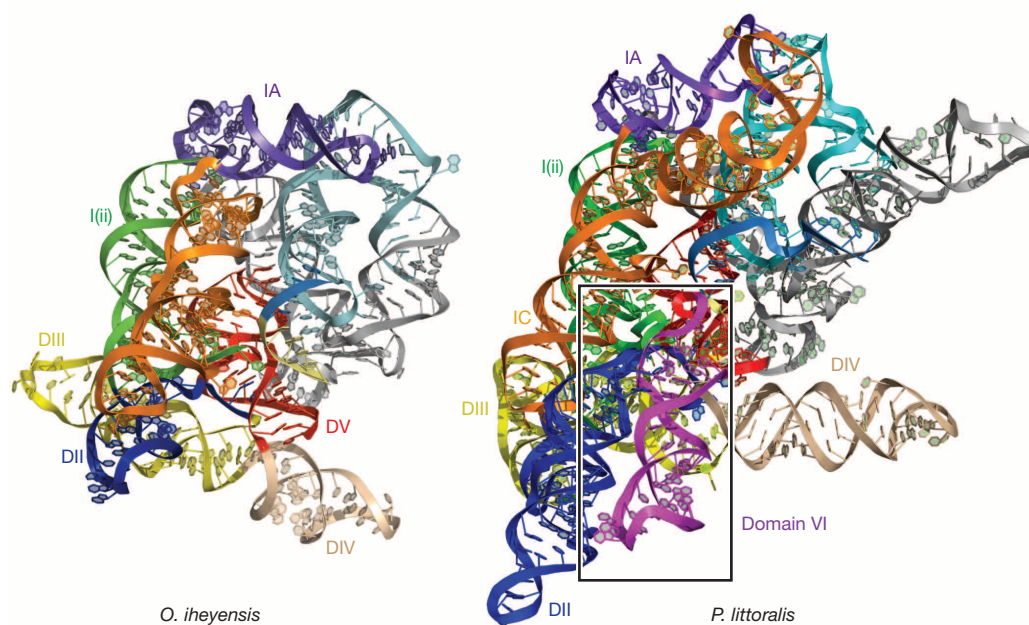


Figure 1 | A comparison of the tertiary structures of *O. iheyensis* and *P. li.LSUI2* group II introns.

P. li.LSUI2 has a significantly larger and more complex structure with a correspondingly greater number of unique RNA tertiary contacts. Domain VI of *P. li.LSUI2* is highlighted within the box.

additional reinforcement. Therefore, DIII functions as an external brace located on the outside surface to stabilize the entirety of the structure. This important role is consistent with deletion or mutagenesis of this

domain rendering the intron unstable and not competent for efficient catalysis¹⁹.

DII positions DVI in the active site

In the *O. iheyensis* structure, DII was truncated to a small stem loop structure, and in many previous biochemical studies of the *al5γ* intron, DII was similarly shortened to study the first step of splicing. We can now visualize the intact DII substructure and find that it serves as a central hub for four different tetraloop receptor interactions (Fig. 3a). DII makes contacts with domains I, III and VI to organize a large portion of the intron structure.

DII has a 'Y-shaped' RNA secondary structure with two stems, D2a and D2b (Extended Data Fig. 1), coaxially stacking on top of each other such that a tetraloop receptor from D2a and a GCAA tetraloop from D2b are facing the same side (Fig. 3a). This combination provides a binding interface for DVI, which contains both a GAGA tetraloop and a tetraloop receptor. Therefore, DVI is tightly placed in the core of the intron via two tandem tetraloop receptor interactions with DII. The interaction between D2a and DVI is known as the η - η' contact²⁰ and we are designating the newly discovered interaction between D2b and DVI as π - π' . The π - π' interaction is especially interesting due to its proximity to the bulged adenosine residue A615, which is the nucleophile for the first step of splicing. The π tetraloop interacts with nucleotides directly adjacent to the bulged adenosine (Fig. 3a) and therefore probably has important effects upon the positioning of this nucleotide within the active site.

Mutagenesis of the GNRA tetraloops to UUCG was done to test the effects of these two interactions on splicing. Disrupting either contact significantly inhibited the second step of splicing, leading to an accumulation of lariat 3' exon and 5' exon (Fig. 3b and Extended Data Fig. 6). However, disrupting both interactions simultaneously resulted in a near complete block of the second step. This indicates that η - η' and

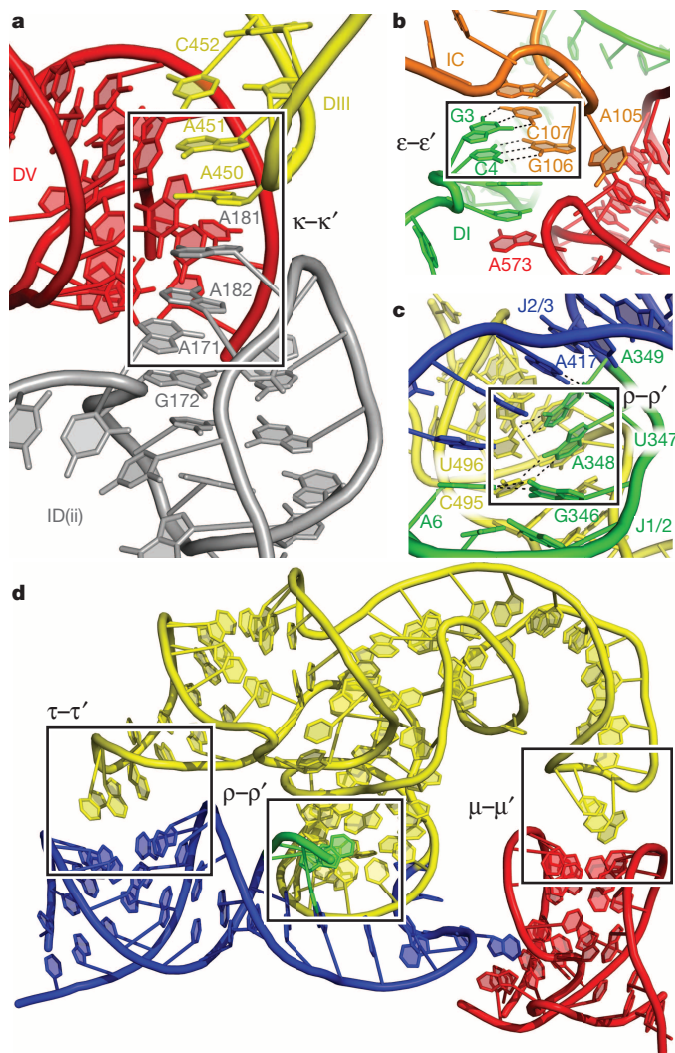


Figure 2 | Tertiary interactions in a IIB intron. **a**, The κ loop (grey) and DIII (yellow) converge to form an extended base stack involving five adenosine residues inserting into the minor groove of DV (red). **b**, G106 and C107 form Watson-Crick pairs with C4 and G3 from the 5' end to form the ϵ - ϵ' interaction that positions the 5' end in the active site. **c**, J1/2 (green) interacts with the 5' end (A6 residue in green), DIII (yellow) and J2/3 (blue). **d**, DIII (yellow) acts as a brace on the surface of the intron and forms three tertiary interactions (boxed) with DI (green), DII (blue) and DV (red).

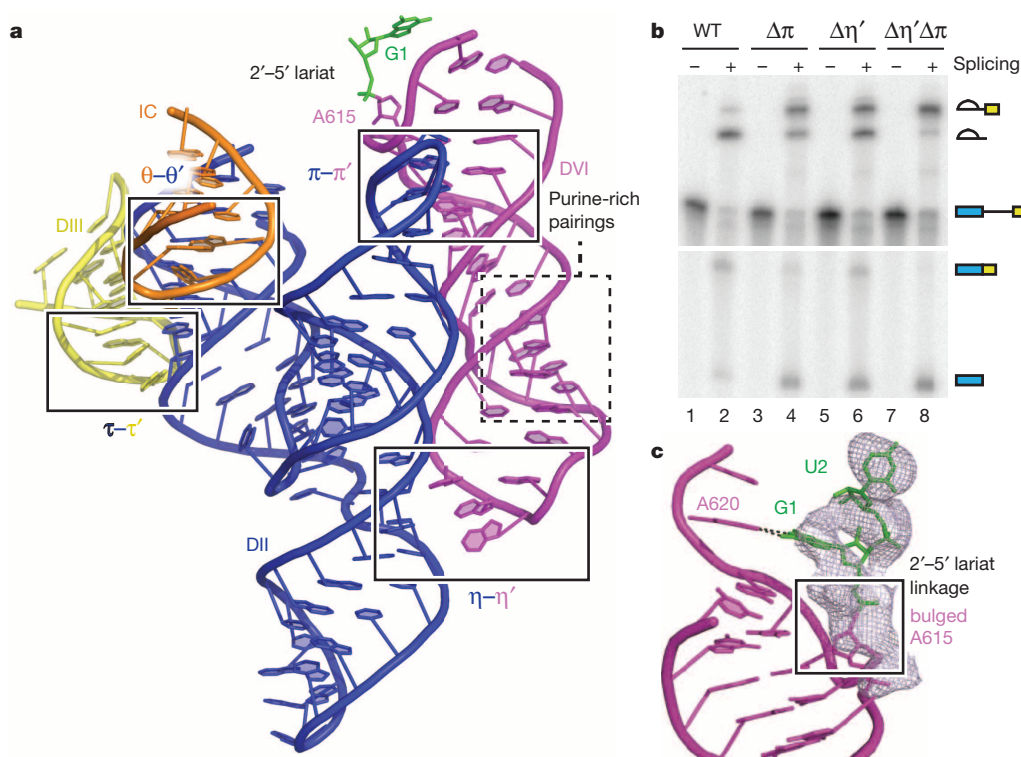


Figure 3 | The position of DVI within the intron structure. **a**, DII forms an inverted Y-shaped structure that engages in four distinct tetraloop-receptor interactions with the IC stem (orange), DIII (yellow) and DVI (purple). DVI interacts with DII via two tandem tetraloop-receptor interactions, η - η' and π - π' . **b**, *In vitro* self-splicing assays of the *P.li.LSUI2* intron. The wild-type (WT) intron efficiently catalyses both steps of splicing and forms intron lariat and ligated exons. Mutagenesis of either η - η' ($\Delta\eta'$)²⁰ or π - π' ($\Delta\pi$) inhibits

π - π' are synergistic interactions essential for the transition to the second step of splicing and probably function through moving the 3' splice site (which is attached to DVI) into the active site.

We can now visualize the overall fold of DVI embedded within a catalytically active intron (Fig. 3a). DVI contains a purine-rich internal loop, which forms non-canonical pairings that induce a slight bend in the helix, allowing DVI to simultaneously form both the η - η' and π - π' contacts with DII. These interactions position the ribose sugar of the bulged A615 directly under the 5' end of the intron. The adenosine in this 2'-5' lariat linkage is highly constrained, since it is connected to phosphates on three different sides of the nucleotide (Extended Data Fig. 7a). The nucleobase component of the A615 residue is disordered (Fig. 3c and Extended Data Fig. 7b), which is consistent with it not having a role in the later stages of splicing. In the current post-catalytic state, the lariat phosphodiester bond is located ~ 20 Å from the active site. Therefore, it has undergone a large-scale movement away from the catalytic core after the first step of splicing. In close proximity to the lariat bond, the 5' and 3' ends of the intron interact with each other through G1 forming a non-canonical base pair with A620 (Fig. 3c), which is important for the second step (Extended Data Fig. 6).

Active site metal ion configuration

Mg^{2+} ions are an absolute requirement for the catalysis of RNA splicing. To identify active site metal ions, soaks were performed using the anomalous scatterer Yb^{3+} , which exhibits the same octahedral coordination geometry as Mg^{2+} and preferentially binds to sites containing highly coordinated magnesium ions^{9,21}. This revealed four large anomalous peaks in the ribozyme core (Fig. 4a). Two of these peaks (M1 and M2) are also found in *O. iheyensis*⁹ and are embedded within DV to coordinate to the junction phosphate between the ligated exons, while the other two peaks (M3 and M4) are coordinated to the 5' end of the intron.

the second step, resulting in the accumulation of lariat 3' exon and 5' exon. A combination of both mutations ($\Delta\eta'\Delta\pi$) nearly blocks the second step of splicing with predominantly lariat 3' exon present. **c**, $F_o - F_c$ density for the 2'-5' lariat phosphodiester bond contoured at 3σ . The nucleobase of A615 is disordered and not visualized. This map was calculated using a model deleted for A615, G1 and U2 (shown in stick format) to avoid model bias. The 5' (G1) and 3' (A620) ends form a non-canonical base pair.

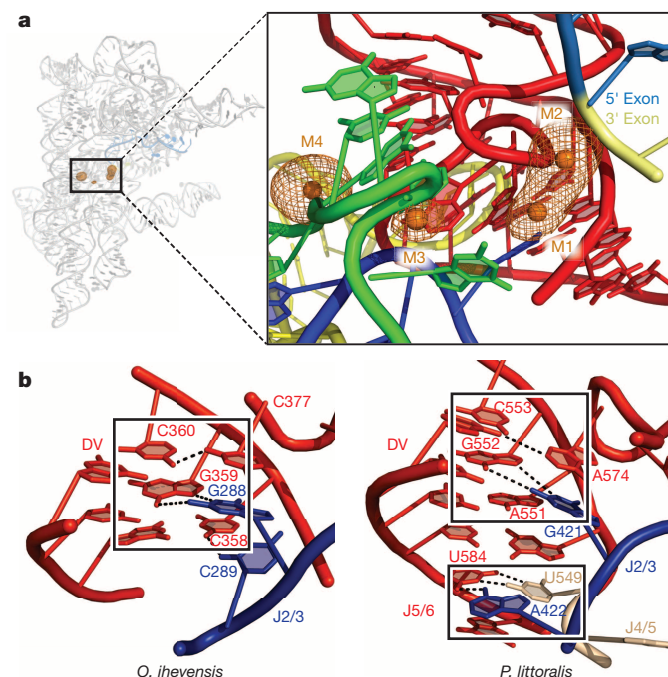


Figure 4 | The core of the of *P.li.LSUI2* intron. **a**, Yb^{3+} anomalous map (orange mesh) contoured at 12σ reveals four large peaks, which correspond to highly coordinated magnesium ions (orange spheres). **b**, Comparison of the catalytic triplexes found in the *O. iheyensis* and *P.li.LSUI2* introns. *O. iheyensis* contains a continuous triple helix spanning the DV catalytic triad (CGC) and J2/3. In *P.li.LSUI2*, the J2/3 residue A422 is disengaged from the catalytic triad (AGC) and forms a base triple with J4/5 and J5/6 nucleotides (U549 and U584, respectively).

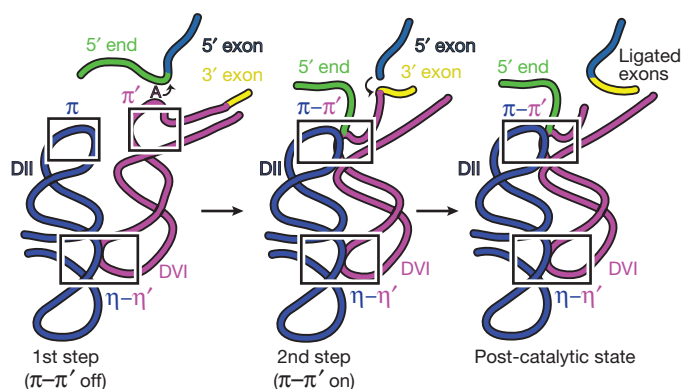


Figure 5 | Model for DVI as the conformational switch for splicing. π - π' mediates the transition between the two steps of catalysis. See text for details.

M3 is located in a binding pocket formed by the highly conserved 5' end (Extended Data Fig. 8a), which has the sequence GUGCG. Nucleotides in this region exhibit a highly contorted backbone configuration that wraps around all sides of M3 due to the ϵ - ϵ' interaction involving residues G3 and C4 (Fig. 2b). Therefore, we postulate that the primary function of ϵ - ϵ' is to order a crucial metal-binding platform which structures the 5' end of the intron.

M4 coordinates to conserved IIB intron residues A6, C7 and A341 to stabilize the 5' end further (Extended Data Fig. 8b). These residues are in close spatial proximity to the recently proposed ι motif (Extended Data Fig. 1) that is known to promote lariat formation and is predicted to serve as a receptor for positioning DVI (ref. 22). An ι G79A point mutant was crystallized, revealing strong signals for M1 and M2; however, M3 and M4 were no longer visible (Extended Data Fig. 8c, d). Splicing assays of this mutant also show a negative effect on the first step of splicing (Extended Data Table 1b). Furthermore, the ϵ - ϵ' interaction, which forms the metal-binding platform for M3 and M4, persists throughout group II intron catalysis¹⁷. Taken together, the data suggest that M3 and M4 participate in the first step of splicing by positioning the 5' splice site in the active site to present the scissile phosphate to M1, M2 and the bulged adenosine. However, it is possible that M3 exists only in the post-catalytic state to stabilize the repositioned lariat bond. Given the rarity of highly coordinated magnesium sites in large RNAs²³, the existence of four such metal ions in close proximity in the active site of *P.li.LSUI2* is striking. In addition, we observe a conserved monovalent ion near the M1/M2 catalytic centre (Extended Data Fig. 8e).

Catalytic triplex rearrangement

In the *O. iheyensis* structure, J2/3 and residues from DV form a catalytic triplex structure essential for the catalysis of splicing and that is also found in the spliceosome²⁴. In the *O. iheyensis* triplex, J2/3 residues G288 and C289 (analogous to G421 and A422 in *P.li.LSUI2*) form base triples with the first two nucleotides of the catalytic triad. Unexpectedly, we observe in the *P.li.LSUI2* structure that one of the J2/3 residues is completely disengaged from the catalytic triad. Specifically, A422 has moved away from the triad and stacks directly underneath the base of the DV helix to form a base triple with residues from the J4/5 and J5/6 linkers (Fig. 4b). The possibility of J2/3 participating in conformational changes has been previously postulated²⁵. However, we can now see the specific nature of this 'switch' with the disengagement of J2/3 from the catalytic triad into an alternate configuration. The linkers between domains are highly conserved for each subclass of group II introns, and based on the *P.li.LSUI2* structure we hypothesize that they are dynamic and modulate splicing.

π - π' is a dynamic interaction

DVI is proposed to engage in large-scale conformational changes between the two steps of splicing^{20,22}. To test this model, a catalytic triad mutant

(AGC→GAU) inactive for splicing was crystallized and solved at 7 Å (Extended Data Fig. 9). Strong electron density was observed for the η - η' interaction between DII and DVI, indicating that this contact persists throughout both steps of splicing and that there is no large-scale change in DVI position. Therefore, a new model is required to explain how DVI mediates the transition between the two steps of splicing.

During the first step, the bulged adenosine must be in close proximity to M1, M2 and the 5' splice site to engage in nucleophilic attack. However, π - π' places the bulged adenosine ~20 Å from the active site in the post-catalytic structure. Mutagenesis of π - π' also has no significant effect on the first step of splicing (Fig. 3b), and is unlikely to be engaged for lariat formation. Furthermore, DII is likely to remain largely stationary during catalysis due to the strong anchoring effect of multiple tetraloop receptor interactions with this domain. On the basis of these observations, we propose that π - π' is a dynamic interaction that toggles DVI between two different states to mediate the transition between the first and second steps of splicing (Fig. 5). In the first step, the bulged adenosine is engaged in the active site for nucleophilic attack at the 5' splice site. At this stage, π - π' exists in the 'off' state where DII is disengaged from the base of DVI. The DVI helix would also presumably exist in a relaxed conformation due to the lack of constraint provided by π - π' . Following lariat formation, DVI probably engages in remodelling of its central internal loop adjacent to A615, causing helical compression to turn 'on' the π - π' interaction, thus sequestering the bulged adenosine away from the active site. A second possible model is that the base pairs between the G₆ sequence (residues 588 to 593) and a pyrimidine-rich tract (612 to 614; 616 to 618) at the proximal side of the DVI stem rearrange to reposition the lariat phosphate and engage π - π' . Both models serve to empty the active site of the 5' end and allow entry of the 3' splice site, which is directly attached to the end of DVI. In fact, the primary function of the lariat may be to covalently attach to the 5' end to provide an attachment point for this pulling action.

Evolutionary implications

This bulged adenosine of DVI is analogous to the branch site adenosine in spliceosomal introns. The branch site sequence UACUAAC (nucleophilic adenosine underlined) pairs to the U2 snRNA to extrude the adenosine from the helix²⁶ as in DVI. Owing to the mechanistic and evolutionary similarities with group II introns, it is likely that the branch site adenosine residue and the 5' end of nuclear introns will adopt a similar spatial arrangement relative to the catalytic core of the spliceosome. We predict that the spliceosomal branch site adenosine will be sequestered after the first step of splicing with an RNA and/or protein contact analogous to the π - π' interaction in *P.li.LSUI2*. In fact, an interaction observed between a region just downstream of the branch site and the U5 snRNA loop may be the spliceosomal counterpart of π - π' (ref. 27).

The *P.li.LSUI2* structure provides a rationale for the phylogenetic conservation of the 5' end. Group II introns have the consensus sequence GUGYG (where Y = U or C)¹¹, which is similar to the GUAYG and GURAG (where R = G or A) sequences found at the 5' ends of yeast and mammalian introns, respectively²⁸. In the spliceosome, the U6 snRNA pairs with the 5' end²⁹ in an analogous manner to the ϵ - ϵ' interaction in group II introns. Therefore, these conserved sequences probably form a similar metal-binding platform in spliceosomal introns, with the 5' end forming a highly distorted backbone to coordinate magnesium ions that orient the splice site in the core.

In regards to the evolutionary rationale for the 2'-5' phosphodiester bond, it is known that the lariat is required for full reversibility of group II intron splicing³⁰. These reverse splicing reactions are the mechanism through which group II introns are able to invade DNA³¹ and disperse throughout genomes. We hypothesize that the lariat bond pre-organizes the core of the intron structure to facilitate this reversibility. Reverse splicing has also been proposed as a vehicle for the prolific expansion of nuclear introns³², which comprise ~25% of mammalian genomes. There is biochemical support for this hypothesis in that the spliceosome has recently

been shown to catalyse reverse splicing reactions³³. The lariat probably has a similar role in spliceosomal introns by allowing reverse splicing to occur, thus accounting for its phylogenetic conservation across the kingdoms by enabling 'selfish introns' to replicate. Therefore, the *P.li.LSUI2* crystal structure has provided the first glimpse of the branched lariat linkage that was probably crucial for intron proliferation in eukaryotes.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 April; accepted 22 August 2014.

Published online 24 September 2014.

- Grabowski, P. J., Padgett, R. A. & Sharp, P. A. Messenger RNA splicing *in vitro*: an excised intervening sequence and a potential intermediate. *Cell* **37**, 415–427 (1984).
- Padgett, R. A., Konarska, M. M., Grabowski, P. J., Hardy, S. F. & Sharp, P. A. Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors. *Science* **225**, 898–903 (1984).
- Konarska, M. M., Grabowski, P. J., Padgett, R. A. & Sharp, P. A. Characterization of the branch site in lariat RNAs produced by splicing of mRNA precursors. *Nature* **313**, 552–557 (1985).
- Peebles, C. L. *et al.* A self-splicing RNA excises an intron lariat. *Cell* **44**, 213–223 (1986).
- van der Veen, R. *et al.* Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing *in vitro*. *Cell* **44**, 225–234 (1986).
- Di Leo, E. *et al.* A point mutation in the lariat branch point of intron 6 of NPC1 as the cause of abnormal pre-mRNA splicing in Niemann-Pick type C disease. *Hum. Mutat.* **24**, 440 (2004).
- Galej, W. P., Oubridge, C., Newman, A. J. & Nagai, K. Crystal structure of Prp8 reveals active site cavity of the spliceosome. *Nature* **493**, 638–643 (2013).
- Fica, S. M. *et al.* RNA catalyses nuclear pre-mRNA splicing. *Nature* **503**, 229–234 (2013).
- Toor, N., Keating, K. S., Taylor, S. D. & Pyle, A. M. Crystal structure of a self-spliced group II intron. *Science* **320**, 77–82 (2008).
- Lambowitz, A. M. & Zimmerly, S. Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb. Perspect. Biol.* **3**, a003616 (2011).
- Toor, N., Hausner, G. & Zimmerly, S. Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. *RNA* **7**, 1142–1152 (2001).
- Michel, F., Umeson, K. & Ozeki, H. Comparative and functional anatomy of group II catalytic introns—a review. *Gene* **82**, 5–30 (1989).
- Costa, M., Fontaine, J. M., Loiseaux-de Goër, S. & Michel, F. A group II self-splicing intron from the brown alga *Pylaiella littoralis* is active at unusually low magnesium concentrations and forms populations of molecules with a uniform conformation. *J. Mol. Biol.* **274**, 353–364 (1997).
- Rest, J. S. & Mindell, D. P. Retroids in archaea: phylogeny and lateral origins. *Mol. Biol. Evol.* **20**, 1134–1142 (2003).
- Toor, N., Robart, A. R., Christianson, J. & Zimmerly, S. Self-splicing of a group IIC intron: 5' exon recognition and alternative 5' splicing events implicate the stem-loop motif of a transcriptional terminator. *Nucleic Acids Res.* **34**, 6461–6471 (2006).
- Boudvillain, M. & Pyle, A. M. Defining functional groups, core structural features and inter-domain tertiary contacts essential for group II intron self-splicing: a NAIM analysis. *EMBO J.* **17**, 7091–7104 (1998).
- Jacquier, A. & Michel, F. Base-pairing interactions involving the 5' and 3'-terminal nucleotides of group II self-splicing introns. *J. Mol. Biol.* **213**, 437–447 (1990).
- Fedorova, O. & Pyle, A. M. A conserved element that stabilizes the group II intron active site. *RNA* **14**, 1048–1056 (2008).
- Fedorova, O., Mitros, T. & Pyle, A. M. Domains 2 and 3 interact to form critical elements of the group II intron active site. *J. Mol. Biol.* **330**, 197–209 (2003).
- Chanfreau, G. & Jacquier, A. An RNA conformational change between the two chemical steps of group II self-splicing. *EMBO J.* **15**, 3466–3476 (1996).
- Adams, P. L., Stahley, M. R., Kosek, A. B., Wang, J. & Strobel, S. A. Crystal structure of a self-splicing group I intron with both exons. *Nature* **430**, 45–50 (2004).
- Li, C. F., Costa, M. & Michel, F. Linking the branchpoint helix to a newly found receptor allows lariat formation by a group II intron. *EMBO J.* **30**, 3040–3051 (2011).
- Klein, D. J., Moore, P. B. & Steitz, T. A. The contribution of metal ions to the structural stability of the large ribosomal subunit. *RNA* **10**, 1366–1379 (2004).
- Fica, S. M., Mefford, M. A., Piccirilli, J. A. & Staley, J. P. Evidence for a group II intron-like catalytic triplex in the spliceosome. *Nature Struct. Mol. Biol.* **21**, 464–471 (2014).
- Marcia, M. & Pyle, A. M. Visualizing group II intron catalysis through the stages of splicing. *Cell* **151**, 497–507 (2012).
- Query, C. C., Moore, M. J. & Sharp, P. A. Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes Dev.* **8**, 587–597 (1994).
- Anokhina, M. *et al.* RNA structure analysis of human spliceosomes reveals a compact 3D arrangement of snRNAs at the catalytic core. *EMBO J.* **32**, 2804–2818 (2013).
- Parker, R. & Guthrie, C. A point mutation in the conserved hexanucleotide at a yeast 5' splice junction uncouples recognition, cleavage, and ligation. *Cell* **41**, 107–118 (1985).
- Lesser, C. F. & Guthrie, C. Mutations in U6 snRNA that alter splice site specificity: implications for the active site. *Science* **262**, 1982–1988 (1993).
- Roitzsch, M. & Pyle, A. M. The linear form of a group II intron catalyzes efficient autocatalytic reverse splicing, establishing a potential for mobility. *RNA* **15**, 473–482 (2009).
- Yang, J., Zimmerly, S., Perlman, P. S. & Lambowitz, A. M. Efficient integration of an intron RNA into double-stranded DNA by reverse splicing. *Nature* **381**, 332–335 (1996).
- Lynch, M. & Richardson, A. O. The evolution of spliceosomal introns. *Curr. Opin. Genet. Dev.* **12**, 701–710 (2002).
- Tseng, C. K. & Cheng, S. C. Both catalytic steps of nuclear pre-mRNA splicing are reversible. *Science* **320**, 1782–1784 (2008).

Acknowledgements We thank S. Banerjee and the staff of the NE-CAT beamlines at the Advanced Photon Source (APS) of Argonne National Laboratory. We thank P. Ghosh, S. Joseph, G. Ghosh, R. Doolittle, Y. Tor, D. Donoghue and T. Wiryaman for comments on the manuscript. We thank R. Das and F.-C. Chou for assistance with phenix.erasser for structure refinement and G. Bricogne for advice on Buster refinement. We also thank N. T. Schirle for preliminary biochemical characterization of the *P.li.LSUI2* intron. R.T.C. was supported by the Cell, Molecular, and Genetics Training Program funded by NIH predoctoral training grant 5T32GM007240. J.K.P. was supported by the UCSF Molecular Biophysics Training Program funded by NIH predoctoral training grant 5T32GM008326. NE-CAT is supported by NIH grant 8P41GM103403-10 and APS is supported by the US DOE under contract number DE-AC02-06CH11357. This work was supported by a Hellman Foundation Fellowship and NIH grant 5R01GM102216 awarded to N.T.

Author Contributions A.R.R. and J.K.P. performed the experiments. A.R.R., R.T.C., J.K.P. and N.T. designed the experiments. A.R.R., R.T.C., J.K.P., K.R.R. and N.T. analysed the data. A.R.R., R.T.C. and N.T. wrote the manuscript with input from all authors.

Author Information Coordinates and structure factors have been deposited in the Protein Data Bank under accession code 4ROD. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.T. (ntoor@ucsf.edu).

METHODS

Cloning and preparation of *Pylaiella littoralis* *P.li.LSUI2* intron RNA. DNA was synthesized (Genscript) corresponding to the second intron interrupting the large ribosomal subunit (LSU) in the mitochondria of the brown alga *Pylaiella littoralis* (*P.li.LSUI2*). The crystallization construct contains a 15-nucleotide 5' exon and a 5-nucleotide 3' exon followed by a HindIII restriction site. The DIV open reading frame (ORF) was removed from the *P.li.LSUI2* intron and replaced with a UUCG tetraloop. This was cloned into the EcoRV site of pUC57. Plasmid was linearized by HindIII digestion before *in vitro* transcription. Non-conserved loops in domains 1, 2 and 4 were changed from the wild-type sequence. Most significantly, changing the sequence UAUUUAUA to UCGACAUAAGG in the ID2 stem loop improved both crystallization rate and diffraction. The final construct retained wild-type splicing activity. Transcription was performed overnight at 37 °C using T7 RNA polymerase in 25 mM MgCl₂, 2 mM spermidine, 5 mM DTT, 40 mM Tris-HCl pH 7.5, 0.05% Triton X-100, 2.5 mM of each NTP, and thermostable inorganic pyrophosphatase (New England Biolabs). CaCl₂ was added to a final concentration of 1.2 mM, treated with DNase I for 45 min, followed by proteinase K digestion for 1 h. The intron reacted to completion during *in vitro* transcription and was subjected to a native purification procedure previously used in the structure determination of the *O. iheyensis* intron⁹. Spliced intron RNA was repeatedly washed with 10 mM MgCl₂ and 5 mM sodium cacodylate pH 6.5 and concentrated to 10 mg ml⁻¹ using a 100 kDa molecular mass cutoff Amicon Ultra-15 column.

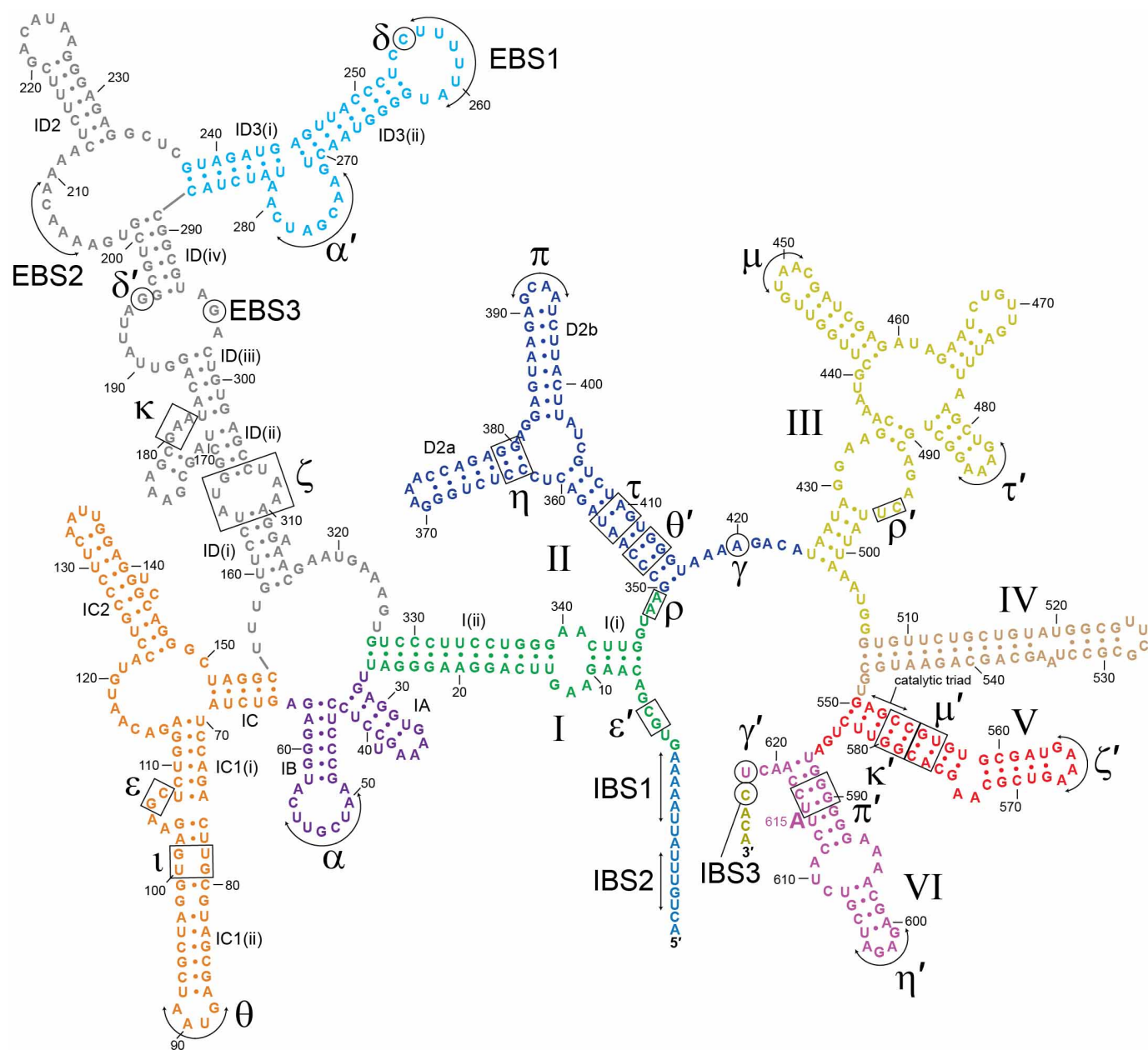
Crystallization. The native crystals were grown in sitting drops by vapour diffusion at 30 °C. Equal volumes of RNA (10 mg ml⁻¹) were mixed with 0.4 mM spermine, 21% 2-methyl-2,4-pentanediol (MPD), 175 mM magnesium acetate tetrahydrate and 90 mM MES monohydrate (pH 5.6). Rod-like crystals appeared within 2 days, and grew to a maximum size of 50 × 50 × 900 µm. Crystals were gradually exchanged into 21% MPD, 100 mM magnesium acetate tetrahydrate, 50 mM MES monohydrate (pH 5.6), 3 mM iridium hexammine, 0.5 mM spermine and 100 mM NaCl, followed by flash freezing in liquid nitrogen. Iridium hexammine was used as an additive for the native crystals as it reduced mosaicity. Pre-catalytic intron RNA was obtained by mutating the AGC catalytic triad of DV to GAU. The mutation maintained stem pairing but completely inhibited splicing activity. Crystallization of this mutant was done by microseeding under conditions described above. The 1 G79A mutant crystallized in 0.2 M ammonium acetate, 10 mM calcium chloride, 50 mM sodium cacodylate pH 6.5 and 10% w/v polyethylene glycol 4000. Resulting crystals were cryoprotected through gradual transfer to the crystallization solution supplemented with 30% ethylene glycol.

Structure determination. The crystal structure of the *P.li.LSUI2* intron was solved at 3.7 Å resolution using multi-wavelength anomalous dispersion (MAD) with crystals soaked in 0.5 mM ytterbium (III) chloride. Yb³⁺ soaks were performed in 21% MPD, 100 mM magnesium acetate tetrahydrate, 50 mM MES monohydrate (pH 5.6), 0.5 mM spermine, 100 mM NaCl and 0.5 mM YbCl₃ for 3 h at room temperature. Tl⁺ soaks were performed in the same manner with 10 mM thallium acetate. X-ray data sets were collected at NE-CAT's 24-ID-C beamline at the Advanced Photon Source (Argonne National Laboratory, Argonne, Illinois). Data was processed

using HKL-2000³⁴, heavy metal sites were identified with SHELXD³⁵ and phasing done using Phenix³⁶ and SHELXE³⁷. RNA nucleotides were modelled using COOT³⁸ and the RCrane plugin³⁹. The phylogenetically predicted secondary structure (Extended Data Fig. 1) guided modelling into the electron density. Structure refinement was done using Buster⁴⁰, Phenix³⁶, DEN⁴¹ and Phenix.Erasser⁴². The 2'-5' phosphodiester bond was restrained using a cif restraint file in phenix.refine. All software was compiled by SBGrid⁴³.

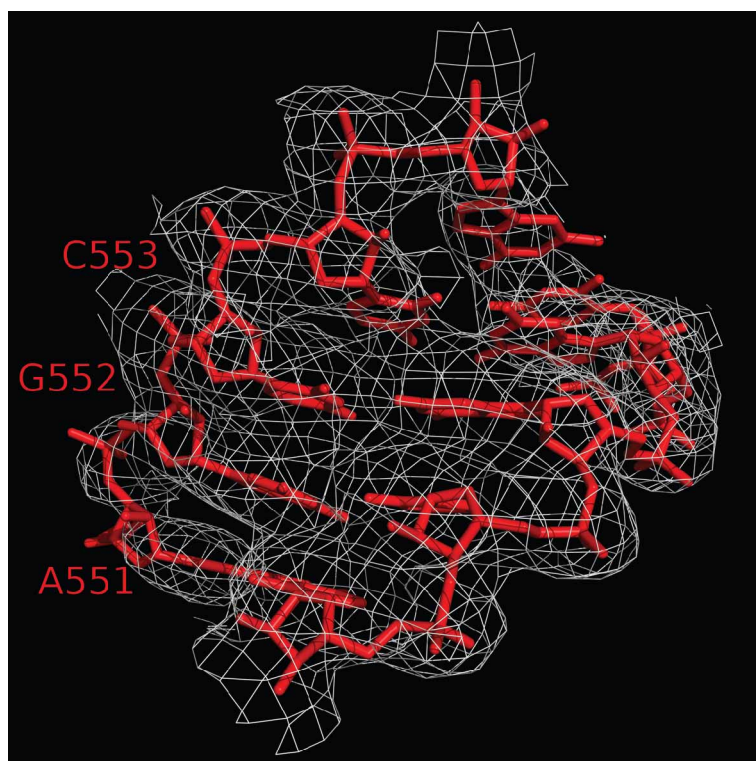
***In vitro* self-splicing assays.** The construct used for the *in vitro* self-splicing assays contained wild-type *P.li.LSUI2* sequence with DIV ORF removed and a 250-nucleotide 5' exon and 75-nucleotide 3' exon. This was cloned into the pUC57 plasmid. Plasmid was linearized using HindIII and used for *in vitro* transcription with T7 RNA polymerase. Radiolabelled transcripts were prepared as above using 10 µCi [α -³²P]UTP (3,000 Ci mmol⁻¹), 0.5 mM UTP, 1 mM other NTPs, and 10 mM MgCl₂. Transcripts were gel purified on a 4% polyacrylamide (19:1)/8 M urea gel, RNA was recovered by diffusion into 300 mM NaCl, 0.01% SDS, 1 mM EDTA. Self-splicing experiments were performed for 30 min at 45 °C in a splicing buffer containing 10 mM MgCl₂, 1 M NH₄Cl, 40 mM Tris-HCl (pH 7.5) and 0.02% SDS. Reactions were stopped by addition of EDTA to a final concentration of 20 mM. Splicing products were resolved using a denaturing 4% polyacrylamide (19:1)/8 M urea gels. Rate constants for the wild-type and G79A mutant were derived from curves fit to a biphasic exponential equation. All splicing assays were done in triplicate.

34. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
35. Schneider, T. R. & Sheldrick, G. M. Substructure solution with SHELXD. *Acta Crystallogr. D* **58**, 1772–1779 (2002).
36. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
37. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008).
38. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
39. Keating, K. S. & Pyle, A. M. RCrane: semi-automated RNA model building. *Acta Crystallogr. D* **68**, 985–995 (2012).
40. Blanc, E. *et al.* Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D* **60**, 2210–2221 (2004).
41. Schroder, G. F., Levitt, M. & Brunger, A. T. Super-resolution biomolecular crystallography with low-resolution data. *Nature* **464**, 1218–1222 (2010).
42. Chou, F. C., Sipakdeevong, P., Dibrov, S. M., Hermann, T. & Das, R. Correcting pervasive errors in RNA crystallography through enumerative structure prediction. *Nature Methods* **10**, 74–76 (2013).
43. Morin, A. *et al.* Collaboration gets the most out of software. *eLife* **2**, e01456 (2013).
44. Chanfreau, G. & Jacquier, A. Interaction of intronic boundaries is required for the second splicing step efficiency of a group II intron. *EMBO J.* **12**, 5173–5180 (1993).
45. Parker, R. & Siliciano, P. G. Evidence for an essential non-Watson-Crick interaction between the first and last nucleotides of a nuclear pre-mRNA intron. *Nature* **361**, 660–662 (1993).
46. Basu, S. *et al.* A specific monovalent metal ion integral to the AA platform of the RNA tetraloop receptor. *Nature Struct. Biol.* **5**, 986–992 (1998).

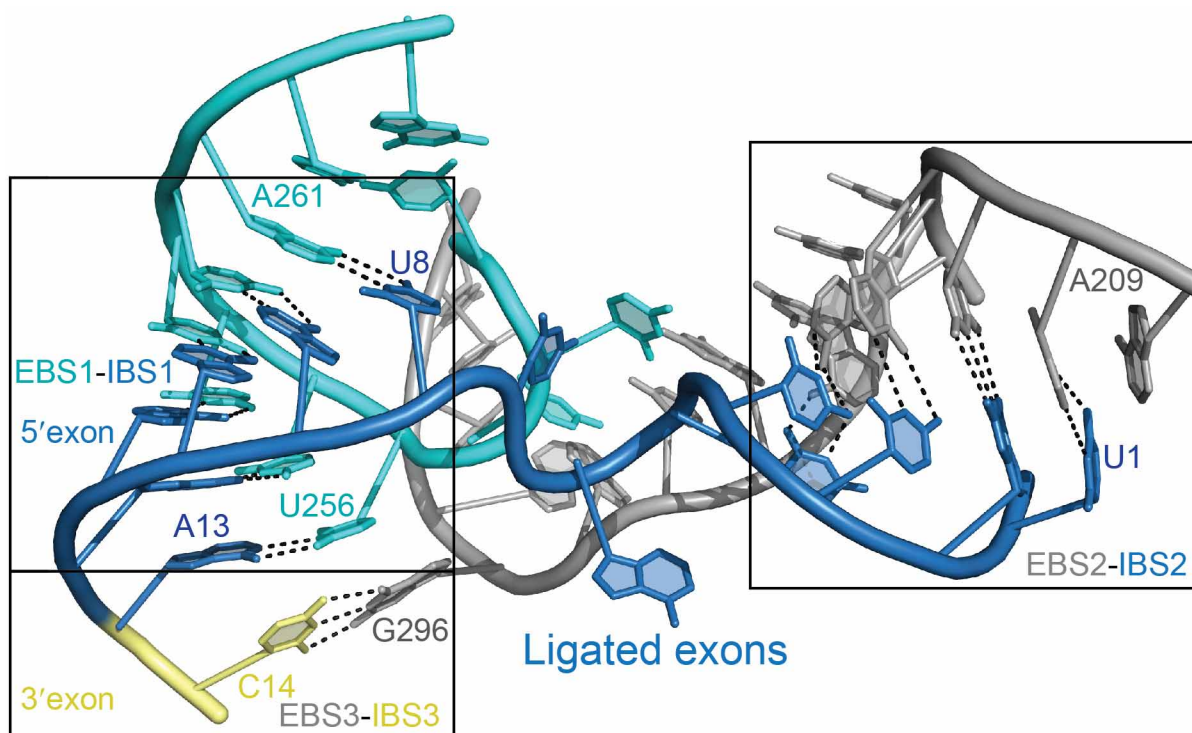


Extended Data Figure 1 | Secondary structure of *P.li.LSUI2* intron crystallization construct. Tertiary interactions are indicated with Greek letters and domains are labelled with Roman numerals. Colouring of the individual

domains is consistent with the overall view of the tertiary structure shown in Fig. 1.

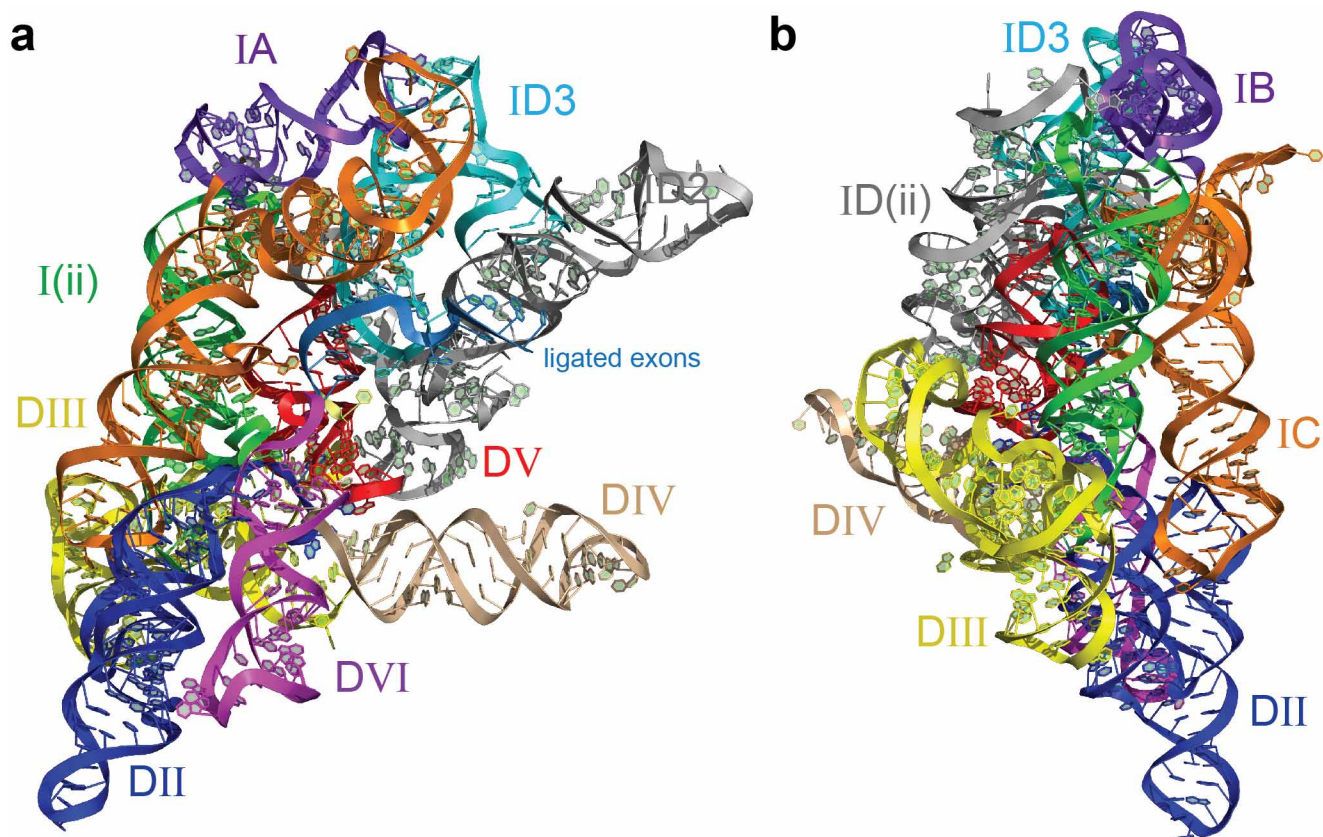


Extended Data Figure 2 | The Yb-MAD experimental, density-modified map of the portion of DV containing the catalytic triad contoured at 1.8σ .



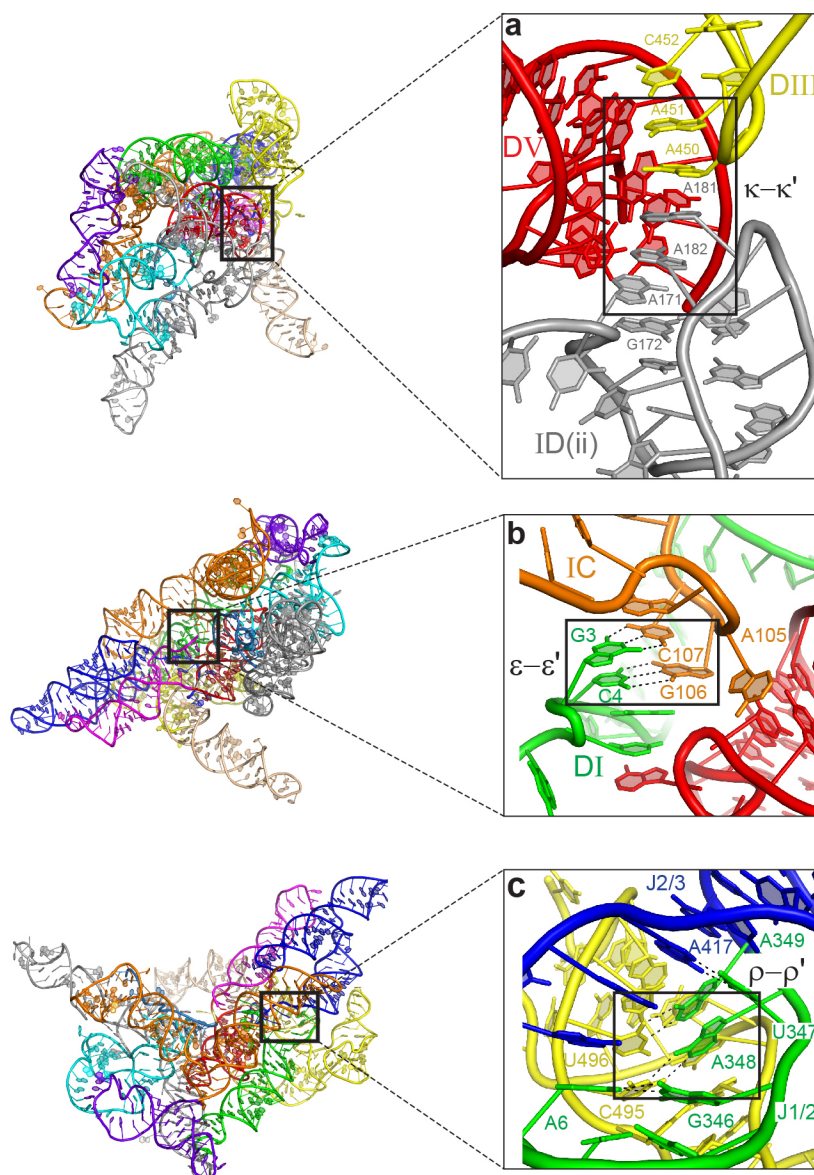
Extended Data Figure 3 | The path of the 5' exon through the intron structure. The EBS1-IBS1 and EBS2-IBS2 interactions position the 5' exon. They do not form a continuous binding interface with the presence of a highly distorted backbone at the junction between these two motifs. As a result, the

helical axes of the EBS1-IBS1 and EBS2-IBS2 pairings are positioned $\sim 90^\circ$ relative to each other. The EBS3-IBS3 interaction places the 3' exon in the active site.

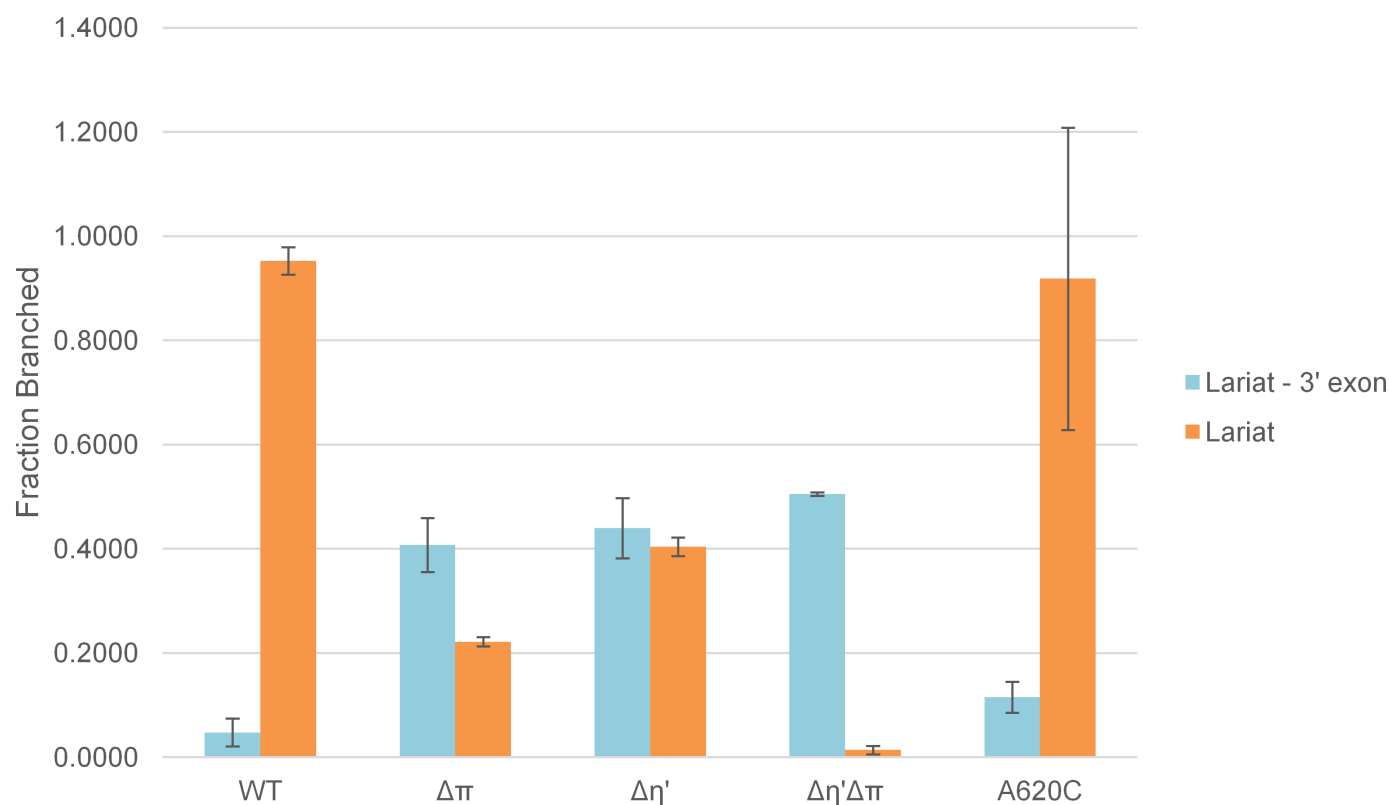


Extended Data Figure 4 | Overall tertiary structure of the *P.li.LSUI2* intron. Individual domains and subdomains are depicted in different colours. Domain

names are labelled with Roman numerals. **a** and **b** show different rotations of the intron structure.

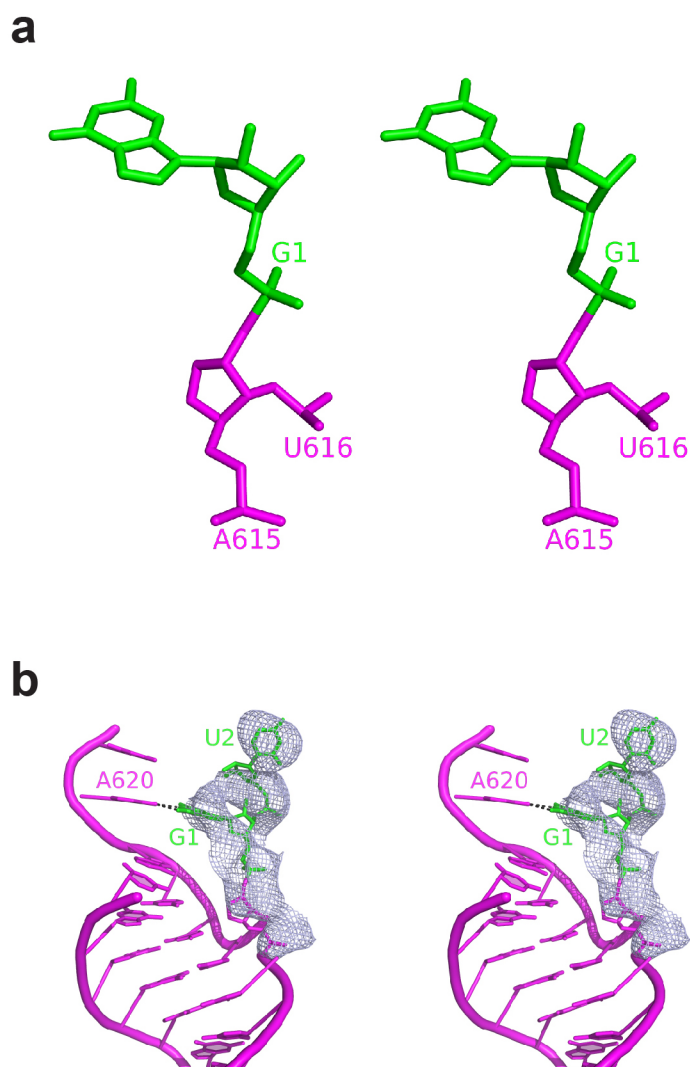


Extended Data Figure 5 | Companion to Fig. 2 showing the location of the individual tertiary interactions relative to the overall structure. **a**, κ - κ' . **b**, ε - ε' . **c**, ρ - ρ' .

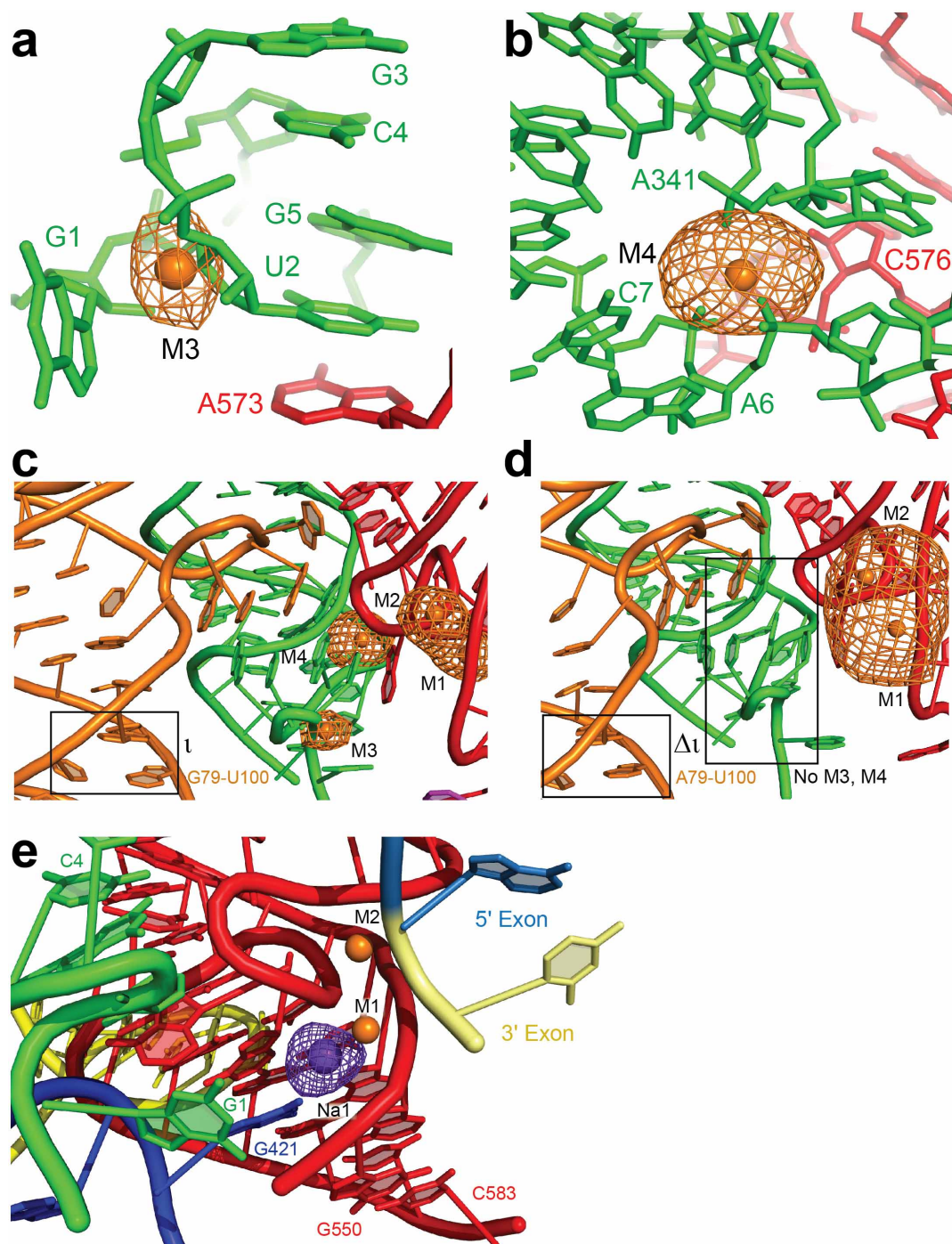


Extended Data Figure 6 | Splicing assays for the DVI mutants showing the proportion of branched product. Blue and orange bars correspond to lariat 3' exon and lariat, respectively. The $\Delta\eta'$ and $\Delta\pi$ mutants accumulate large amounts of lariat 3' exon, thus indicating a second-step splicing defect. The $\Delta\eta'\Delta\pi$ double mutant is almost completely blocked before the second step. The A620C mutant shows 2.4-fold greater accumulation of lariat 3' exon compared to the wild-type intron, indicating that the interaction between G1 and A620 is

important for the second step. In the yeast *al5 γ* intron, G1 instead interacts with the penultimate residue⁴⁴, indicating a certain degree of flexibility for this pairing. There is evidence for a similar interaction between the termini of nuclear introns⁴⁵ involving nearby (but not exactly equivalent) residues, which also has a significant effect upon the second step of splicing. Therefore, the 5' and 3' ends of nuclear introns may have a similar arrangement within the spliceosome.



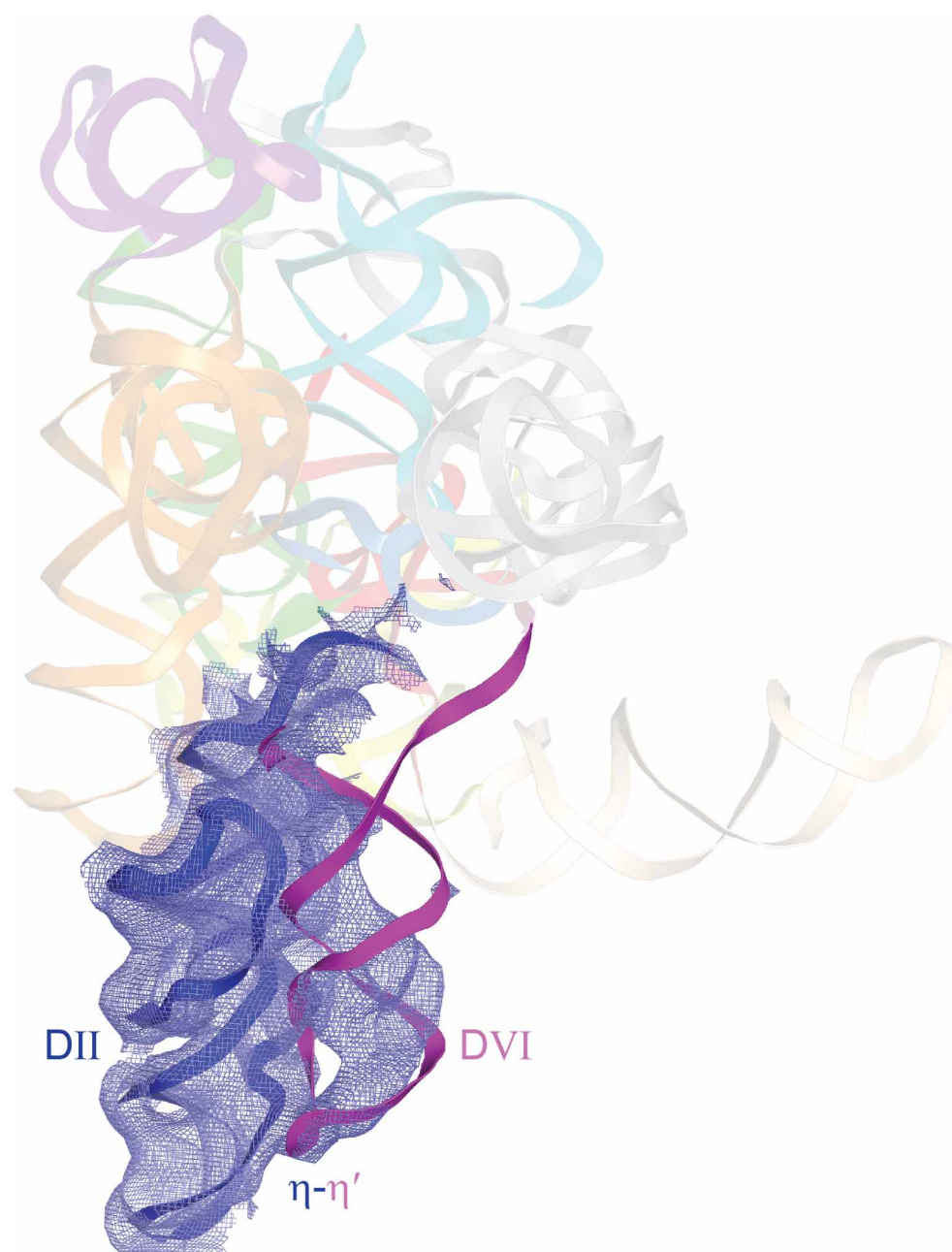
Extended Data Figure 7 | Detailed view of the lariat bond. **a**, The lariat 2'-5' phosphodiester bond in wall-eyed stereo format. **b**, Stereo version of Fig. 3c. See Fig. 3c legend for details. $F_o - F_c$ density for the lariat bond contoured at 3σ .



Extended Data Figure 8 | Anomalous maps identifying core metal ions.

a, b, Depiction of the RNA ligands surrounding metals M3 and M4, respectively. Yb³⁺ anomalous map contoured at 9σ. **c**, Yb³⁺ anomalous map for wild type contoured at 9σ. **d**, Compared with the wild-type intron, the Yb³⁺ anomalous map for the ι G79A mutant (contoured at 4σ) is lacking the peaks corresponding to M3 and M4, even at a lower contour level. **e**, Tl⁺ was used as a

probe for monovalent ions in the RNA structure⁴⁶. The Tl⁺ anomalous map (purple mesh contoured at 5.5σ) revealed a strong peak located 3.8 Å from M1 that coordinates to the nucleobase of J2/3 residue G421 and the backbone of DV nucleotide G550. This sodium ion Na1 (purple sphere) is significantly closer to M1 than the equivalent K⁺ ion found in *O. iheyensis*²⁵. Otherwise, this monovalent ion binding site is relatively conserved between these two introns.



Extended Data Figure 9 | $2F_o - F_c$ density for DVI in the pre-catalytic structure contoured at 1σ . The η - η' interaction persists throughout the splicing reaction and is visible in the pre-catalytic state. The weaker density for the central region of DVI suggests a partially disordered, dynamic region with

possible helical remodelling in the conserved internal loop during splicing. The general pattern of side-by-side packing of domains II and VI persists between the two steps. Catalytic triad mutation consisted of an AGC→GAU substitution.

Extended Data Table 1 | X-ray and kinetic data for *P.li.LSUI2***a** Data collection and refinement statistics.

	<i>P.li.LSUI2</i> native	DV Triad Mutant	G79A Mutant	Yb ³⁺	Tl ⁺
Data collection				Peak	Inflection
Space group	C222 ₁	C222 ₁	C222	C222 ₁	C222 ₁
Cell					
<i>a</i> , <i>b</i> , <i>c</i> (Å)	163.7, 255.4, 136.8	161.9, 264.5, 137.5	211.7, 457.2, 179.5	164.6, 257.7, 138.0	161.2, 257.7, 136.3
α , β , γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Wavelength (eV)	11218.5	12662	8949.5	8949.5	8946.5
Resolution (Å)	150.0-3.68 (3.74-3.68)	50.0-7.25 (7.37-7.25)	50.0-9.75 (9.92-9.75)	50.0-4.65 (4.73-4.65)	50.0-4.49 (4.57-4.49)
<i>R</i> _{sym}	14.9 (>100)	9.5 (77.9)	8.6 (97.0)	8.8 (73.9)	9.2 (72.8)
<i>I</i> / σ <i>I</i>	6.4 (0.6)	13.2 (1.7)	14.3 (1.8)	16.1 (2.1)	21.8 (2.2)
Completeness	99.9 (99.9)	94.6 (82.5)	96.3 (96.9)	99.9 (99.6)	97.9 (88.7)
Redundancy	6.8 (3.8)	5.7 (4.7)	6.5 (6.0)	4.4 (3.7)	3.6 (2.8)
CC*	(0.743)				
Refinement					
Resolution (Å)	81.8-3.68				
No. reflections	31107				
<i>R</i> _{work} / <i>R</i> _{free}	23.9/27.4				
No. atoms	13979				
RNA	13471				
Ligand/ion	393				
Water	115				
B-factors					
RNA	201.6				
Ligand/ion	181.5				
Water	177.0				
R.m.s.					
Bond lengths	0.017				
Bond angles	1.483				

*Highest resolution shell is shown in parenthesis.

b Comparison of wild-type (WT) and G79A splicing rates.

	<i>k</i> _{fast} (min ⁻¹)	<i>k</i> _{slow} (min ⁻¹)
WT	9.51 ± 1.09	0.136 ± 0.0246
G79A	2.97 ± 1.57	0.0767 ± 0.00942

A mass of less than 15 solar masses for the black hole in an ultraluminous X-ray source

C. Motch¹, M. W. Pakull¹, R. Soria², F. Grisé^{1,3,4,5} & G. Pietrzyński^{6,7}

Most ultraluminous X-ray sources¹ have a typical set of properties not seen in Galactic stellar-mass black holes. They have luminosities of more than 3×10^{39} ergs per second, unusually soft X-ray components (with a typical temperature of less than about 0.3 kiloelectronvolts) and a characteristic downturn^{2,3} in their spectra above about 5 kiloelectronvolts. Such puzzling properties have been interpreted either as evidence of intermediate-mass black holes^{4,5} or as emission from stellar-mass black holes accreting above their Eddington limit^{6,7}, analogous to some Galactic black holes at peak luminosity^{8,9}. Recently, a very soft X-ray spectrum was observed in a rare and transient stellar-mass black hole¹⁰. Here we report that the X-ray source P13 in the galaxy NGC 7793¹¹ is in a binary system with a period of about 64 days and exhibits all three canonical properties of ultraluminous sources. By modelling the strong optical and ultraviolet modulations arising from X-ray heating of the B9Ia donor star, we constrain the black hole mass to be less than 15 solar masses. Our results demonstrate that in P13, soft thermal emission and spectral curvature are indeed signatures of supercritical accretion. By analogy, ultraluminous X-ray sources with similar X-ray spectra and luminosities of up to a few times 10^{40} ergs per second can be explained by supercritical accretion onto massive stellar-mass black holes.

We organized an X-ray–ultraviolet–optical spectrophotometric monitoring programme of the ultraluminous X-ray source (ULX) P13¹² from late 2009 until late 2013 using NASA’s Swift and Chandra satellites, the European Space Agency’s XMM-Newton satellite and the European Southern Observatory’s Very Large Telescope (VLT). These data were supplemented by photometry obtained at the Warsaw 1.3 m telescope (Las Campanas Observatory) in 2004 and 2005 and by an archival Chandra observation obtained in 2003. Details of the observations and data reduction as well as further analysis of the results are reported in the Supplementary Information.

Chandra detected P13 in 2003¹³ with a 0.3–10 keV X-ray luminosity of $L_X(0.3\text{--}10\text{ keV}) \approx 4 \times 10^{39} \text{ erg s}^{-1}$. Our 73 d-long Swift X-ray Telescope (XRT) monitoring carried out in 2010 recorded a 0.3–10 keV X-ray luminosity in the range of $(4.8 \pm 0.5) \times 10^{39} \text{ erg s}^{-1}$ down to less than $1.6 \times 10^{38} \text{ erg s}^{-1}$ on one occasion. A similar X-ray luminosity, of $(2.0 \pm 0.1) \times 10^{39} \text{ erg s}^{-1}$, was detected in our last XMM-Newton observation in November 2013. The 2003 Chandra spectrum displays a spectral break at ~ 4.2 keV. Our 2013 XMM-Newton observation confirms the break seen by Chandra and reveals a soft disk-black-body-like component with $k_B T_{\text{in}} \approx 0.3$ keV, where k_B is Boltzmann’s constant and T_{in} is the temperature at the inner disk radius (Extended Data Table 1 and Extended Data Fig. 1). Therefore, P13 exhibits all the hallmarks of a canonical ultraluminous X-ray state^{2,3}. Remarkably, Swift/XRT did not detect P13 in any of the individual pointings performed from August 2011 until June 2013. Stacking Swift faint-state data reveals the source at $L_X(0.3\text{--}10\text{ keV}) = (5 \pm 1) \times 10^{37} \text{ erg s}^{-1}$ (90% confidence level), which is a factor of 100 less than the previously seen bright X-ray state. Scheduled and serendipitous Chandra and XMM-Newton observations

carried out in 2011 and 2012 detected the source at the same low X-ray luminosity and will be reported elsewhere.

Optical spectra suggest a B9I spectral type (Fig. 1). In addition to high-order Balmer absorption lines, the spectrum exhibits Balmer emission up to at least H γ as well as He II $\lambda 4,686$ (that is, at a wavelength of $\lambda = 4,686 \text{ \AA}$) and Bowen C III–N III emission. Assuming that minimum light (visual magnitude, $V = 20.50$ mag; Extended Data Figs 2a–d and 3a) represents stellar light yields an absolute magnitude of only $M_V = -7.50$ mag (distance, $d = 3.7$ Mpc (ref. 14); mean extinction, $A_V = 0.16$ mag (refs 15, 16)). Such a high optical luminosity implies a type Ia supergiant classification with an initial mass of 20–25 solar masses (M_\odot) and a present mass of $18M_\odot$ – $23M_\odot$ (ref. 17). Because the bolometric luminosity is $L_{\text{bol}} \approx 5 \times 10^{38} \text{ erg s}^{-1}$, which is almost one-tenth of the maximum observed X-ray luminosity, we expect X-ray heating effects to be noticeable.

The Las Campanas observations and the Swift Ultraviolet and Optical Telescope (UVOT) u-band and V-band 2010 monitorings all show two consecutive maxima, providing possible hints of a ~ 2 month-long orbital period. The power spectrum analysis of the 2004–2011 V-band light curve (Extended Data Fig. 4) reveals two aliasing periods, one at $P = 65.165$ d and one at $P = 63.340$ d. Corresponding periodicities are found in the He II radial velocities. Importantly, analysing times of visual and ultraviolet photometric maxima over a 8 yr time interval reveals a phase jitter of up to ± 0.09 , which might reflect a superorbital period of ~ 5 – 8.8 yr (Extended Data Figs 5 and 6).

The pattern of radial velocity variations with orbital phase changed very significantly between 2010 (X-ray bright state) and 2011 (X-ray faint state) (Extended Data Fig. 7). Balmer absorption lines show what may be a coherent variation with orbital phase in 2010, and the pattern of variability is clearly more complex in 2011. The total velocity amplitude of the absorption lines is $\sim 160 \text{ km s}^{-1}$.

Interestingly, the shape and amplitude of the optical light curve (Extended Data Fig. 8) and the mean value of the He II $\lambda 4,686$ equivalent width do not seem to depend on the observed X-ray luminosity. The relative amplitude of the u-band (central wavelength, $\lambda_{\text{central}} = 3,465 \text{ \AA}$; light-curve full amplitude, $\Delta u = 1.0$ mag) and V-band ($\lambda_{\text{central}} = 5,500 \text{ \AA}$; $\Delta V = 0.5$ mag) light curves and the behaviour of the $V-I$ colour index suggest that the hemisphere of the supergiant star facing the compact companion experiences a strong X-ray heating effect. An X-ray source with a tenth of the nominal X-ray luminosity would brighten the star by only ~ 0.1 mag in V at maximum light and would have basically no effect in the faint X-ray state. Therefore, we conclude that in 2011 part of the companion star photosphere continued to be illuminated by a luminous X-ray source that is shielded from our view. The Galactic X-ray binary Hercules X-1 exhibits similar bright and faint X-ray states¹⁸, as well as periodic phase shifts of photometric maxima¹⁹. By analogy, we suggest that a tilted, precessing accretion disk is at the origin both of the bright and faint X-ray states and of the phase jitter of optical maximum light.

¹Observatoire astronomique de Strasbourg, Université de Strasbourg, CNRS, UMR 7550, 11 rue de l’Université, F-67000 Strasbourg, France. ²International Centre for Radio Astronomy Research, Curtin University, GPO Box U1987, Perth, Western Australia 6845, Australia. ³Instituto de Astrofísica de Canarias, Calle Vía Láctea, s/n, E-38205 La Laguna, Tenerife, Spain. ⁴Universidad de La Laguna, Departamento de Astrofísica, E-38206 La Laguna, Tenerife, Spain. ⁵Department of Physics and Astronomy, University of Iowa, Van Allen Hall, Iowa City, Iowa 52242, USA. ⁶Universidad de Concepción, Departamento de Astronomía, Casilla 160-C, Concepción, Octava Region, Chile. ⁷Warsaw University Observatory, Aleje Ujazdowskie 4, 00-478 Warszawa, Poland.

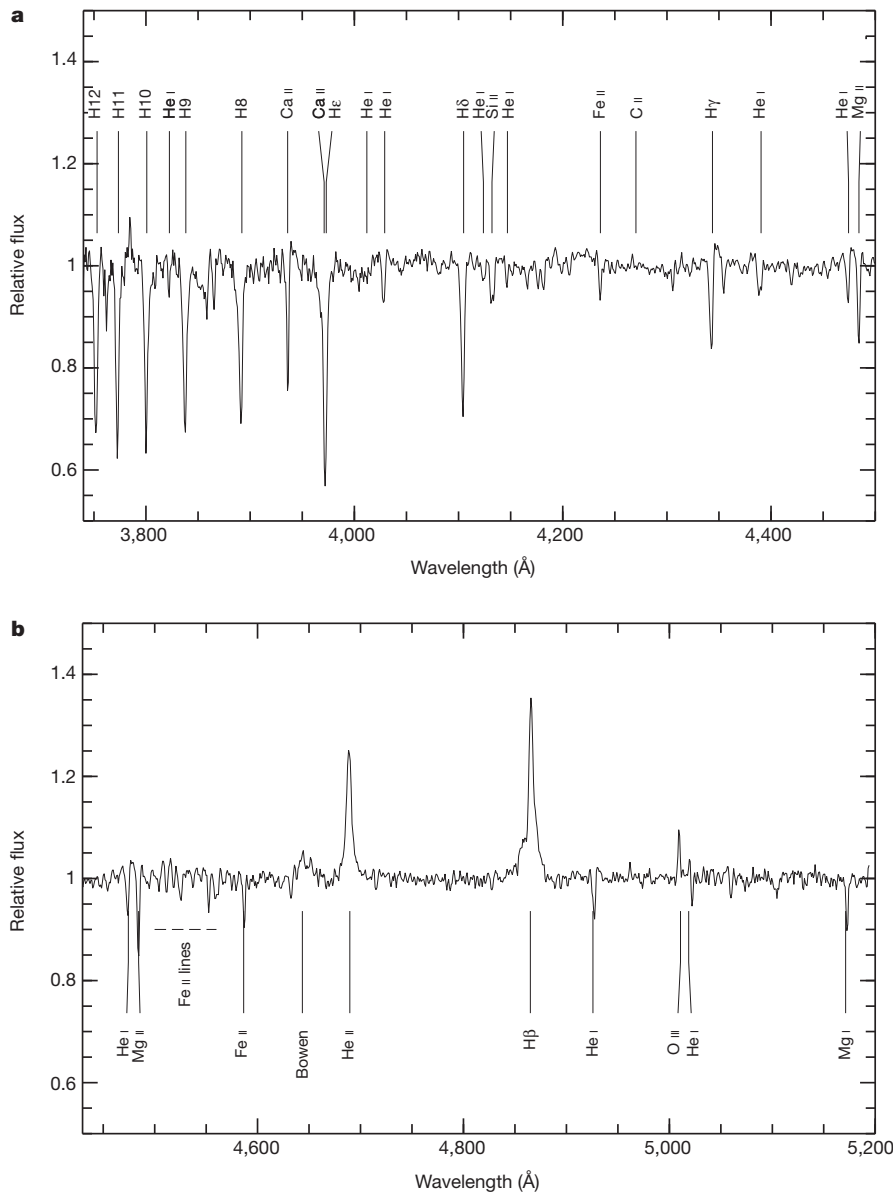


Figure 1 | The mass donor star of P13 has a B9Ia spectral type. The rectified mean of 11 spectra (a, blue part; b, red part) obtained in 2010 and of 11 spectra obtained in 2011 for a total integration time of 15.52 h and corresponding to an average V-band magnitude of $V_{\text{aver}} = 20.31$ mag. All spectra were acquired at the ESO VLT using FORS2. The Mg II $\lambda 4,481$ absorption line is deeper than observed in 2009¹¹ ($V_{\text{aver}} = 20.20$ mag) and is now more consistent with a B9 type²⁵. This indicates that the spectral type of P13 may vary slightly with X-ray illumination, as expected. The strength of the Si II doublet at $\lambda 4,128\text{--}4,130$ (a) compared with neighbouring He I lines also indicates this spectral type. In a similar manner, the appearance of a group of Fe II lines around $\lambda 4,500$ (b) suggests a B9Ia spectral type. A subset of spectra obtained close to minimum light ($V_{\text{aver}} = 20.4$ mag) suggests the same spectral type, indicating that the intrinsic (non-X-ray-illuminated) spectral type of the mass-donor star is indeed B9Ia. Temperature scales for B supergiants^{26,27} imply an effective temperature of $11,000 \pm 1,000$ K.

To constrain the geometry and dynamics of the system, we simultaneously fitted the V-band and UVOT u-band light curves using the eclipsing light curve (ELC) code²⁰. We tested four X-ray luminosity levels, namely 0.7, 1, 1.4 and 2 times a nominal value of $4.2 \times 10^{39} \text{ erg s}^{-1}$ (derived from the diskbb + comptt fit to the Chandra spectrum extrapolated to the 0.3–20 keV range), to account for the observed X-ray variability and for a possible undetected component radiating at energies below or above the observed 0.3–10 keV X-ray range. The ELC model included a dark accretion disk casting X-ray shadows on the X-ray-heated stellar hemisphere. Modelling the optical disk emission caused by X-ray heating, and the absence of veiling in the high-order Balmer absorption lines, indicates that optical light is fully dominated by stellar emission (Extended Data Fig. 9). Examples of fitted light curves are shown in Fig. 2.

We explored the parameter space spanning black-hole/star mass ratios $M_{\text{BH}}/M_{\star} = 0.1\text{--}10$; inclinations from 0° to 90° ; effective stellar temperatures $T_{\text{eff}} = 10,000, 11,000$ and $12,000$ K; the four choices of L_{X} mentioned above; B9Ia star masses of $18M_{\odot}$ and $23M_{\odot}$; and two rotation states of the mass-donor star, namely periastron-synchronized and non-rotating. In each case, the B9Ia star was assumed to fill its Roche lobe at periastron because the total wind mass loss rate of a B9Ia star²¹ is lower or equal to the mass accretion rate required to explain the

bright X-ray state ($\sim 7 \times 10^{-7} M_{\odot} \text{ yr}^{-1}$). The accretion disk had a fixed radius of 0.7 times the Roche lobe radius at periastron. For each of these parameters, we obtained best-fit values for the remaining orbital and disk parameters. We computed the bolometric magnitude M_{bol} by equating the radius of the mass-donor star with that of the corresponding volume-averaged Roche lobe at periastron. We then constrained the range of possible masses of the X-ray source by forcing ELC fits to be acceptable at the 99.7% level, forcing computed bolometric magnitudes to be consistent with observations and forcing possible eclipses to be shorter than the maximum of ~ 7 d allowed by Swift 2010 data. All acceptable orbital solutions implied velocity amplitudes of Balmer absorption lines smaller than the maximum observed range.

All constraints converge towards a black hole mass less than $\sim 15M_{\odot}$, irrespective of the incident X-ray luminosity in the range considered here (Fig. 3). Black holes more massive than $\sim 15M_{\odot}$ imply Roche lobes that are too small at periastron to accommodate the large B9Ia star. All acceptable orbital solutions require a significant eccentricity of $e = 0.27\text{--}0.41$ (Table 1). The full radial velocity amplitude of the compact object varies from 120 to 290 km s^{-1} , a range of values consistent with those observed in 2010 and 2011 for the He II emission line.

The evolved nature of the donor star suggests that mass transfer to the black hole happens on a thermal timescale ($\sim 10^5$ yr) as the supergiant

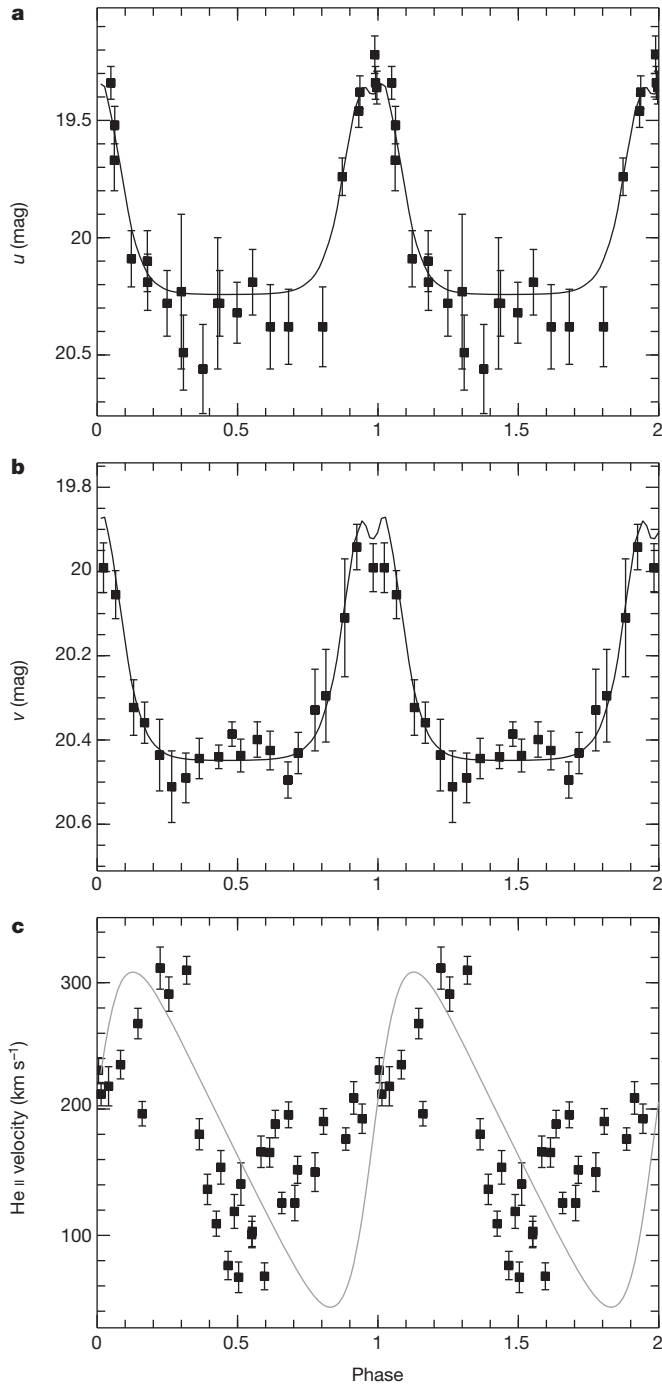


Figure 2 | Examples of V and u light-curve fits. Examples of acceptable models, using the ELC code, obtained by simultaneously fitting V and Swift/UVOT u light curves. Model parameters: mass-donor mass, $23M_{\odot}$; $T_{\text{eff}} = 12,000$ K; $M_{\text{BH}} = 6.9M_{\odot}$; $\log(L_X) = 39.925$; $i = 80^\circ$; $e = 0.33$; periastron angle, 93° . The disk radius has a β opening angle of 6.9° . $\chi^2 = 38.19$ for 39 degrees of freedom. Error bars in all panels show statistical uncertainties at the 1σ level. We assume an orbital period of 63.52 d and a superorbital period of 2,620 d. The spin angular velocity of the donor star is synchronized with its orbital angular velocity at periastron. **a**, **b**, ELC model fits to the u (**a**) and binned V (**b**) light curves. Occultation of the X-ray-heated star hemisphere by the dark disk accounts for the small dip present at maximum light. **c**, Model radial velocity curve of the X-ray source overplotted on observed He II emission line radial velocities. The model curve appears shifted by ~ 0.3 in phase with respect to the observations. This suggests that in P13 the location of the He II $\lambda 4,686$ emitting region does not accurately trace the motion of the compact object, a situation similar to that encountered in several Galactic low-mass X-ray binaries^{28,29}.

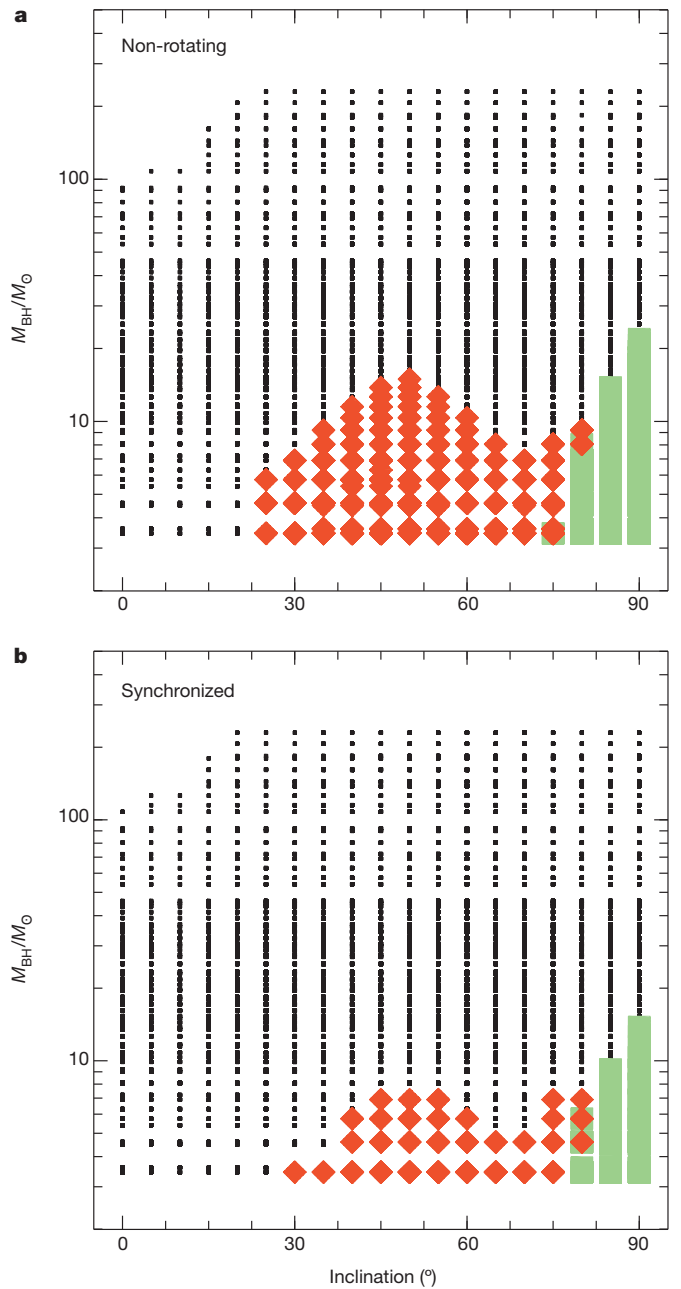


Figure 3 | Constraints on the mass of the compact star. Constraints on the mass of the compact object plotted for $T_{\text{eff}} = 11,000 \pm 1,000$ K (consistent with the B9Ia spectral type of the mass-donor star); $\log(L_X) = 39.475, 39.625, 39.775$ and 39.925 ; B9Ia stellar masses of $18M_{\odot}$ and $23M_{\odot}$, mass ratios M_{BH}/M_* in the range 0.1–10, and mass-donor stars non rotating (**a**) or synchronized with orbital rotation at periastron (**b**). For each mass-donor star mass, system inclination, mass ratio and T_{eff} , the best ELC fit to the u and V light curves provides the eccentricity, the periastron angle and phase, and the radial velocity of the light barycentre of the companion star and of the X-ray source. We show only solutions with $M_{\text{BH}} > 3M_{\odot}$. Small black-filled squares: solutions providing statistically acceptable fits to the light-curves at the 99.7% confidence level. Large symbols: Solutions implying mass-donor optical luminosities compatible with the size of the Roche lobe radius at periastron. Large green-filled squares: excluded solutions implying eclipse durations longer than the maximum allowed by our Swift X-ray monitoring in 2010. Large red diamonds: finally allowed black hole masses considering all possible values of the input parameters. The maximum allowed black hole masses are $\sim 7M_{\odot}$ and $\sim 15M_{\odot}$ for a synchronized mass-donor star and a non-rotating mass-donor star, respectively. These values are obtained for $M_* = 23M_{\odot}$.

Table 1 | Allowed range of orbital parameters

Synchronized	Yes	No
$M_{\text{BH}} (M_{\odot})$	3.45–6.9	3.45–14.95
$\text{Min}(\chi^2)$	0.77	0.73
e	0.27–0.38	0.30–0.41
$\omega (^{\circ})$	79–118	84–124
$i (^{\circ})$	25–80	20–80
$M_{\text{bol}} (\text{mag})$	−7.72 to −7.03	−7.89 to −6.97
$R_{\text{RL}} (R_{\odot})$	64–77	64–82
X-ray heating (mag)	0–0.04	0–0.09
$\text{RV}_2 (\text{km s}^{-1})$	171–285	124–286
$\text{RV}_1 (\text{km s}^{-1})$	18–64	19–116

$\text{Min}(\chi^2)$ is the minimum reduced value (39 degrees of freedom) of all acceptable solutions; ω is the periastron angle; R_{RL} is the radius of the Roche lobe of the mass donor star in solar radii (R_{\odot}); X-ray heating is the additional V-band magnitude due to X-ray heating at minimum light; and RV_1 and RV_2 are the full amplitudes of the radial velocity of the mass-donor star and of the black hole, respectively.

rapidly expands²². This is more than an order of magnitude shorter than its main-sequence lifetime, and implies that supergiant ULXs are much rarer than systems with unevolved mass donors²³. Given the significant eccentricity of the orbit, it is likely that Roche lobe overflow started only after the star began crossing the Hertzsprung gap²⁴.

The intrinsic X-ray luminosity of $\sim 4 \times 10^{39} \text{ erg s}^{-1}$ is about twice the Eddington luminosity of a $15M_{\odot}$ accreting black hole. We thus confirm that P13 is a genuine Eddington or super-Eddington source and that its extreme X-ray luminosity does not reflect the presence of an intermediate-mass black hole. Hence, we do have direct evidence that the characteristic ULX X-ray spectrum, with both a medium energy break and a soft X-ray excess, is the signature of an Eddington or super-Eddington regime.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 April 2013; accepted 28 July 2014.

- Feng, H. & Soria, R. Ultraluminous X-ray sources in the *Chandra* and *XMM-Newton* era. *New Astron. Rev.* **55**, 166–183 (2011).
- Gladstone, J. C., Roberts, T. P. & Done, C. The ultraluminous state. *Mon. Not. R. Astron. Soc.* **397**, 1836–1851 (2009).
- Roberts, T. P., Warwick, R. S., Ward, M. J., Goad, M. R. & Jenkins, L. P. *XMM-Newton*/EPIC observations of the ultraluminous X-ray source NGC 5204 X-1. *Mon. Not. R. Astron. Soc.* **357**, 1363–1369 (2005).
- Colbert, E. J. M. & Mushotzky, R. F. The nature of accreting black holes in nearby galaxy nuclei. *Astrophys. J.* **519**, 89–107 (1999).
- Miller, J. M., Fabbiano, G., Miller, M. C. & Fabian, A. C. X-ray spectroscopic evidence for intermediate-mass black holes: cool accretion disks in two ultraluminous X-ray sources. *Astrophys. J.* **585**, L37–L40 (2003).
- Poutanen, J., Lipunova, G., Fabrika, S., Butkevich, A. G. & Abolmasov, P. Supercritically accreting stellar mass black holes as ultraluminous X-ray sources. *Mon. Not. R. Astron. Soc.* **377**, 1187–1194 (2007).
- Kawashima, T. *et al.* Comptonized photon spectra of supercritical black hole accretion flows with application to ultraluminous X-ray sources. *Astrophys. J.* **752**, 18–29 (2012).
- Done, C., Wardziński, G. & Gierliński, M. GRS 1915+105: the brightest Galactic black hole. *Mon. Not. R. Astron. Soc.* **349**, 393–403 (2004).
- Middleton, M. J. *et al.* Bright radio emission from an ultraluminous stellar-mass microquasar in M 31. *Nature* **493**, 187–190 (2013).
- Liu, J., Bregman, J. N., Bai, Y., Justham, S. & Crowther, P. Puzzling accretion onto a black hole in the ultraluminous X-ray source M 101 ULX-1. *Nature* **503**, 500–503 (2013).
- Motch, C., Pakull, M. W., Grisé, F. & Soria, R. The supergiant optical counterpart of ULX P13 in NGC 7793. *Astron. Nachr.* **332**, 367–370 (2011).
- Read, A. M. & Pietsch, W. ROSAT observations of the Sculptor galaxy NGC 7793. *Astron. Astrophys.* **341**, 8–22 (1999).

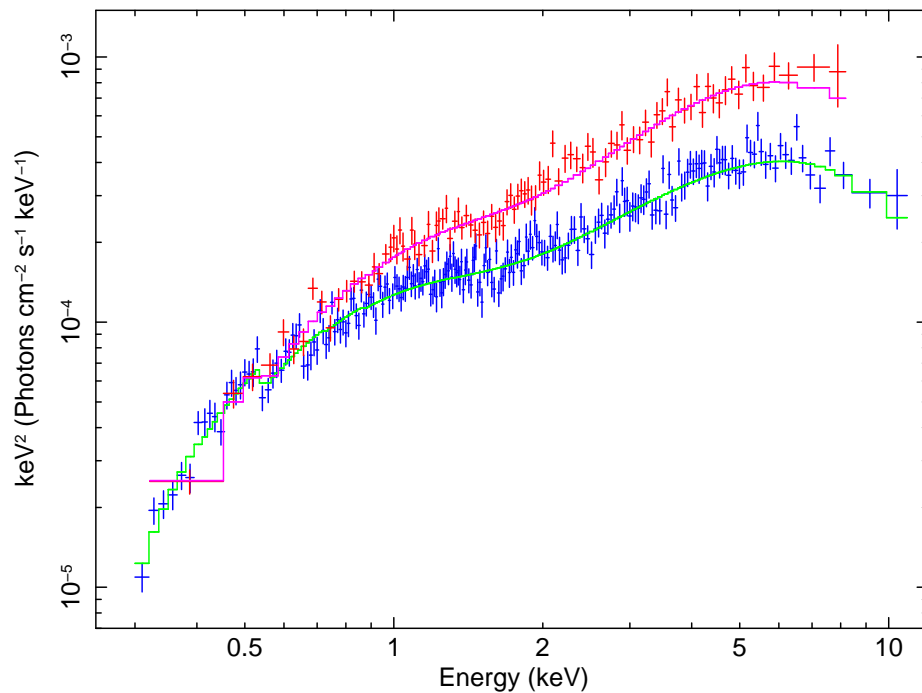
- Pannuti, T. G. *et al.* A *Chandra* observation of the nearby Sculptor group Sd galaxy NGC 7793. *Astron. J.* **142**, 20–38 (2011).
- Radburn-Smith, D. J. *et al.* The GHOSTS survey. I. Hubble Space Telescope Advanced Camera for Surveys data. *Astrophys. J. Suppl. Ser.* **195**, 18–45 (2011).
- Schlegel, D. J., Finkbeiner, D. P. & Davis, M. Maps of dust infrared emission for use in estimation of reddening and cosmic microwave background radiation foregrounds. *Astrophys. J.* **500**, 525–553 (1998).
- Pietrzyński, G. *et al.* The Araucaria Project: first Cepheid distance to the Sculptor group galaxy NGC 7793 from variables discovered in a wide-field imaging survey. *Astron. J.* **140**, 1475 (2010).
- Meynet, G. & Maeder, A. Stellar evolution with rotation. V. Changes in all the outputs of massive star models. *Astron. Astrophys.* **361**, 101–120 (2000).
- Giacconi, R. *et al.* Further X-ray observations of Hercules X-1 from Uhuru. *Astrophys. J.* **184**, 227–236 (1973).
- Deeter, J. *et al.* Analysis of periodic optical variability in the compact X-ray source HER X-1/HZ Herculis. *Astrophys. J.* **206**, 861–868 (1976).
- Orosz, J. A. & Hauschildt, P. H. The use of the NextGen model atmospheres for cool giants in a light curve synthesis code. *Astron. Astrophys.* **364**, 265–281 (2000).
- Kudritzki, R. P. *et al.* The wind momentum-luminosity relationship of galactic A- and B-supergiants. *Astron. Astrophys.* **350**, 970–984 (1999).
- Podsiadlowski, P., Rappaport, S. & Han, Z. On the formation and evolution of black hole binaries. *Mon. Not. R. Astron. Soc.* **341**, 385–404 (2003).
- Rappaport, S. A., Podsiadlowski, P. & Pfahl, E. Stellar-mass black hole binaries as ultraluminous X-ray sources. *Mon. Not. R. Astron. Soc.* **356**, 401–414 (2005).
- Patruno, A. & Zampieri, L. The black hole in NGC 1313 X-2: constraints on the mass from optical observations. *Mon. Not. R. Astron. Soc.* **403**, L69–L73 (2010).
- Lennon, D. J., Dufton, P. L. & Fitzsimmons, A. Galactic B-supergiants. I - an atlas of O9–B9 supergiant spectra from 3950 Å to 4950 Å. *Astron. Astrophys.* **94** (suppl.), 569–586 (1992).
- Markova, N. & Puls, J. Bright OB stars in the Galaxy. IV. Stellar and wind parameters of early to late B supergiants. *Astron. Astrophys.* **478**, 823–842 (2008).
- Zorec, J. *et al.* Fundamental parameters of B supergiants from the BCD system. I. Calibration of the (λ_{11}, D) parameters into T_{eff} . *Astron. Astrophys.* **501**, 297–320 (2009).
- Still, M. D., Quaintrell, H., Roche, P. D. & Reynolds, A. P. Spectral signatures of reprocessing on the companion and accretion disc of Hercules X-1. *Mon. Not. R. Astron. Soc.* **292**, 52–62 (1997).
- Pearson, K. J. *et al.* Multiwavelength observations of EXO 0748–676. II. Emission-line behavior. *Astrophys. J.* **648**, 1169–1180 (2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements R.S. acknowledges an Australian Research Council Discovery Projects funding scheme (project number DP 120102393). F.G. acknowledges support from CNES (CNRS/INSU/CNES contract no. 92532) and partly from the Spanish Ministry of Science and Innovation (MICINN) under grant AYA 2010-18080. Support from the Ideas Plus programme of the Polish Ministry of Science is also acknowledged. We thank J. Orosz for providing us with the ELC code. We acknowledge the Swift team for executing our observing programme. This work is based on observations made with ESO telescopes at the La Silla Paranal Observatory under programme IDs 084.D-0881 and 087.D-0602, and uses observations made with the NASA/ESA Hubble Space Telescope (HST) and obtained from the Hubble Legacy Archive, which is a collaboration between the Space Telescope Science Institute (STScI/NASA), the Space Telescope European Coordinating Facility (ST-ECF/ESA) and the Canadian Astronomy Data Centre (CAD/C/NRC/CSA). The scientific results reported in this article are based in part on observations made by the Chandra X-ray Observatory and in part on observations obtained with XMM-Newton (OBSIDs 0693760101 and 0693760401), an ESA science mission with instruments and contributions directly funded by ESA member states and NASA.

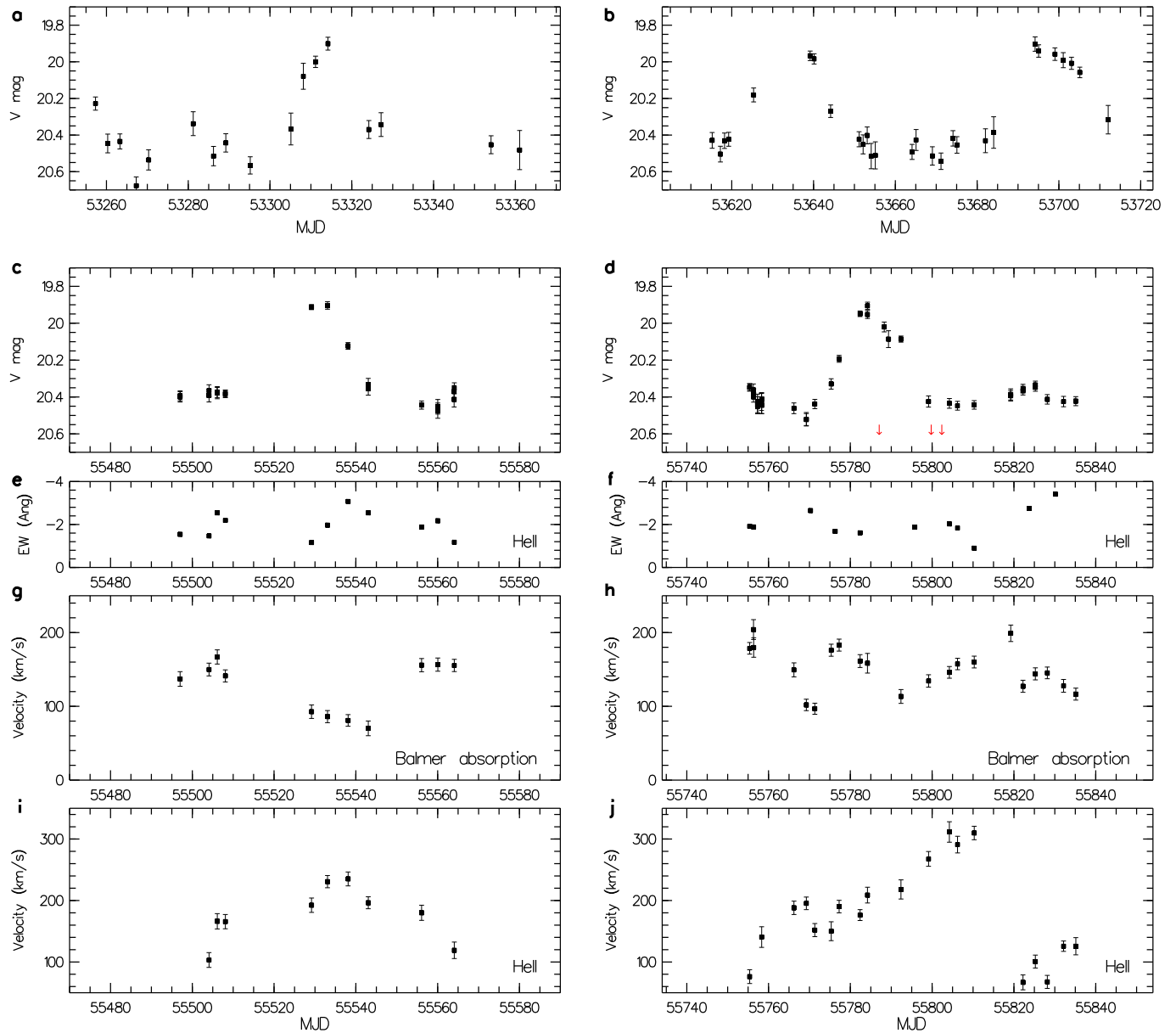
Author Contributions C.M. wrote the manuscript with comments from all authors. M.W.P. identified the optical counterpart of P13 and initiated optical observations at the ESO. C.M. and F.G. analysed the spectroscopic and photometric data from several runs at the ESO Very Large Telescope. F.G. analysed the HST data. G.P. provided photometric data from the Warsaw telescope at the Las Campanas Observatory. R.S., F.G. and C.M. designed and analysed the Chandra, XMM-Newton and Swift X-ray observations. C.M. carried out the light-curve fitting using the ELC code. C.M., M.W.P., R.S. and F.G. made significant contributions to the interpretation and discussion of the data. All authors participated in the review of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.M. (christian.motch@unistra.fr).



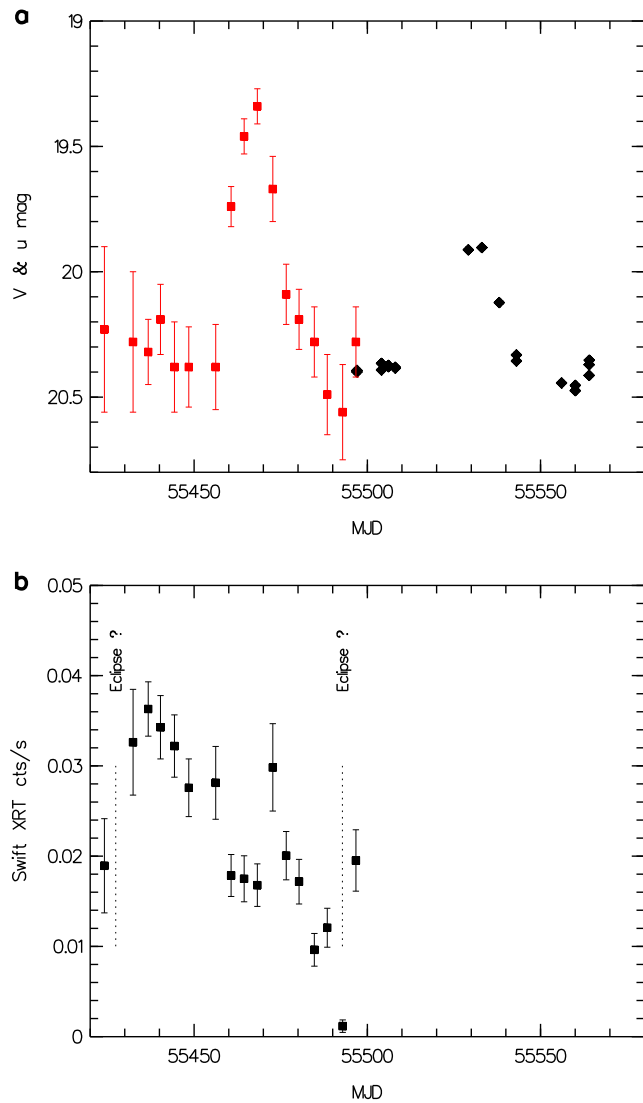
Extended Data Figure 1 | XMM-Newton and Chandra bright-state spectra. The $E \times f(E)$ unfolded energy distribution of the 2003 Chandra/ACIS-S (upper (red) data points) and of the 2013 XMM-Newton/EPIC combined spectra (lower (blue) data points) fitted with the diskir model. Best-fit model

parameters are listed in Extended Data Table 1. Spectra shown here have been re-binned with a minimum signal-to-noise ratio of 8 in each bin for display purposes. Error bars show the 1σ statistical error.

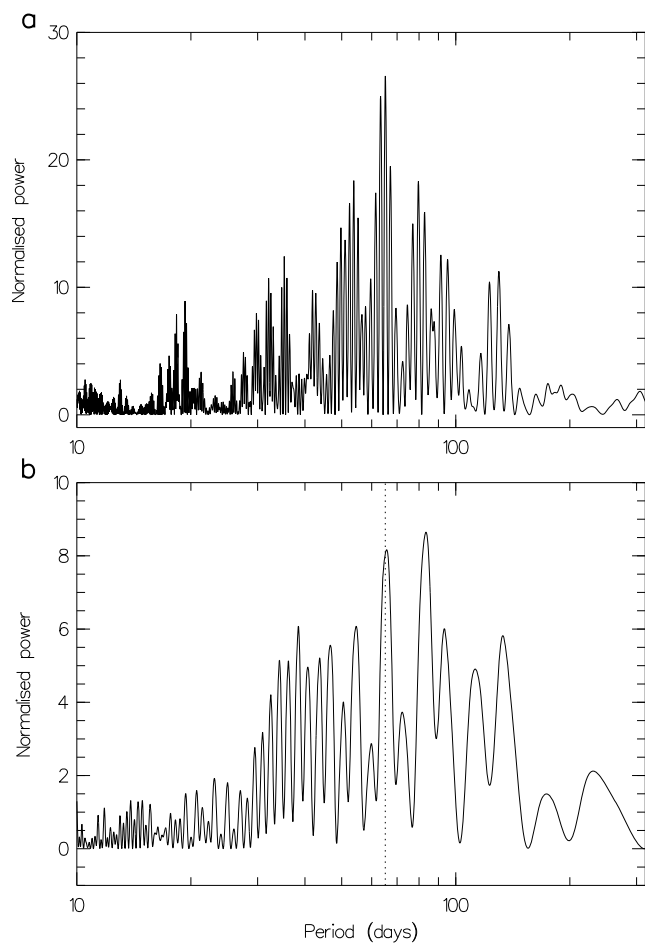


Extended Data Figure 2 | Optical data. Optical light and radial velocity curves of P13. **a, b**, 2004 and 2005 Las Campanas photometry; **c, d**, ESO VLT light curves; **e, f**, equivalent width of the He II $\lambda 4,686$ emission line; **g–j**, radial velocity curves of Balmer absorption lines (**g, h**) and the He II emission line (**i, j**). P13 was X-ray bright in 2010 (**c, e, g, i**) and X-ray faint

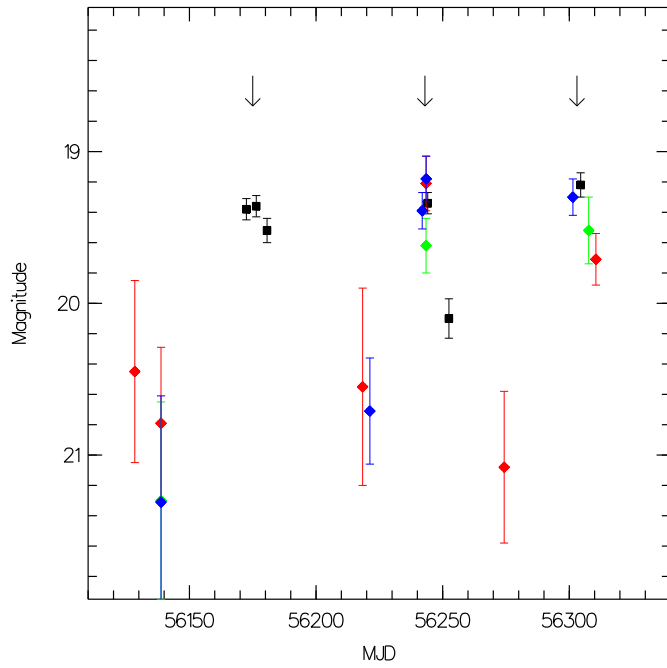
in 2011 (**d, f, h, j**). Red arrows in **d** mark times of the Chandra (near optical maximum) and Swift observations which detected P13 in the faint-X-ray state. Error bars in all panels show statistical uncertainties at the 1σ level. MJD, modified Julian date.



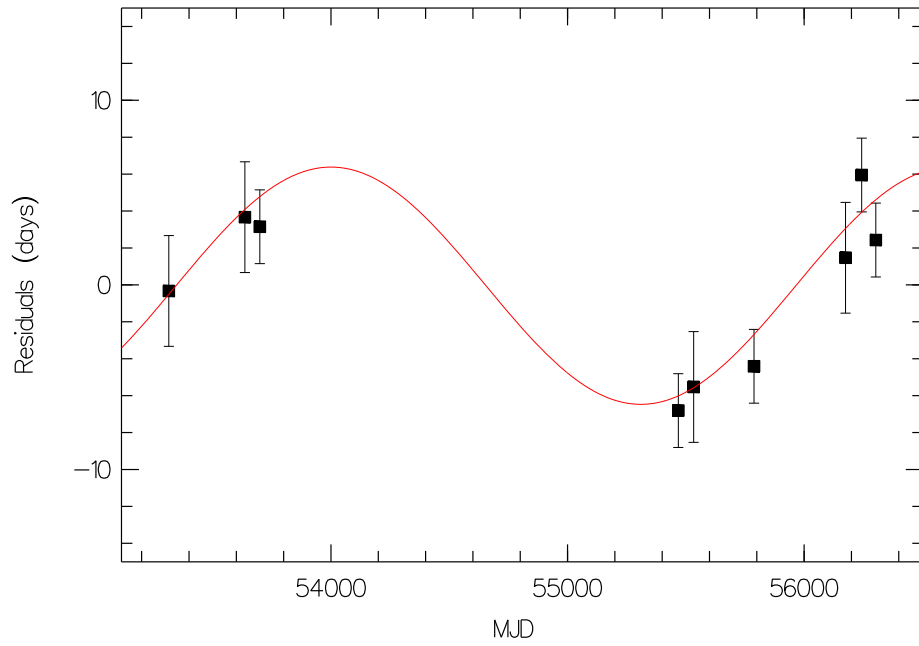
Extended Data Figure 3 | The 2010 observations. **a**, Red squares, UVOT u band ($\lambda_{\text{eff}} = 3,470 \text{ \AA}$); black lozenges, ESO VLT V light curve. Note the successive maxima separated by $\sim 64 \text{ d}$, more pronounced in the u band ($\Delta u \approx 1.0 \text{ mag}$) than in V ($\Delta V \approx 0.5 \text{ mag}$). **b**, Swift/XRT 0.3–10 keV light curve obtained from 2010 August 16 to 2010 October 27. The count-to-flux factor was computed by fitting the XMM-Newton diskir model to the average bright-state spectrum. This implies unabsorbed (0.3–10 keV) X-ray luminosities in the range of $4.8 \times 10^{39} \text{ erg s}^{-1}$ down to less than $1.6 \times 10^{38} \text{ erg s}^{-1}$ on MJD 55492.9 with a weighted average of $(2.8 \pm 0.2) \times 10^{39} \text{ erg s}^{-1}$. Error bars in all panels show statistical uncertainties at the 1σ level. Errors in V (not shown) are typically lower than 0.04 mag.



Extended Data Figure 4 | Power spectra. **a**, Lomb-Scargle power spectrum of the entire V light curve. **b**, Lomb-Scargle power spectrum of the He II radial velocity curve. The dotted line shows the position of the highest peak of the V-band periodogram ($P = 65.165$ d).

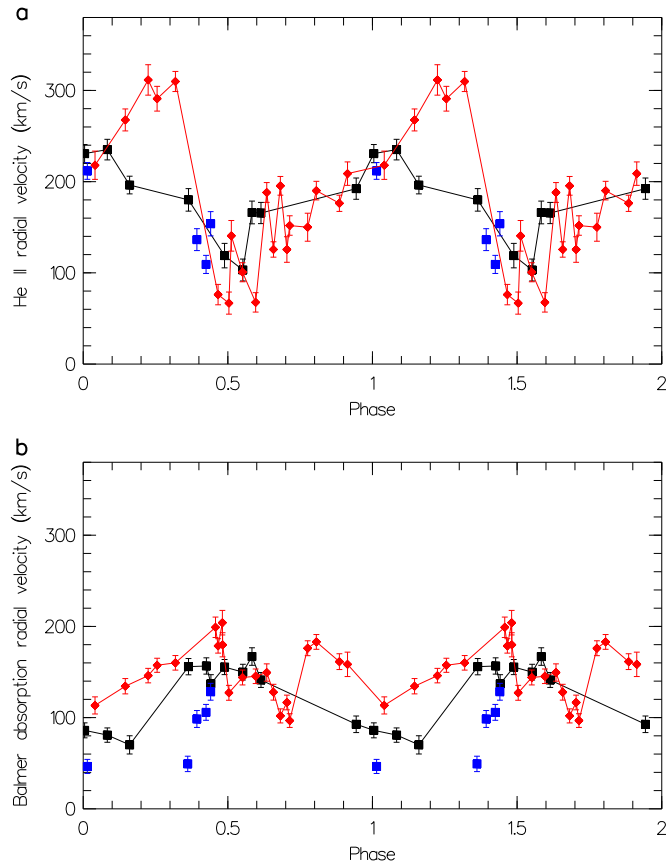


Extended Data Figure 5 | UVOT multiband photometric light curve. Times of photometric maxima used to constrain the superorbital period in 2012 and 2013 are shown with arrows. Black squares, *u* ($\lambda_{\text{central}} = 3,465 \text{ \AA}$); green lozenges, *uw1* ($\lambda_{\text{central}} = 2,600 \text{ \AA}$); red lozenges, *um2* ($\lambda_{\text{central}} = 2,246 \text{ \AA}$); blue lozenges, *uw2* ($\lambda_{\text{central}} = 1,928 \text{ \AA}$). Error bars show statistical uncertainties at the 1σ level.

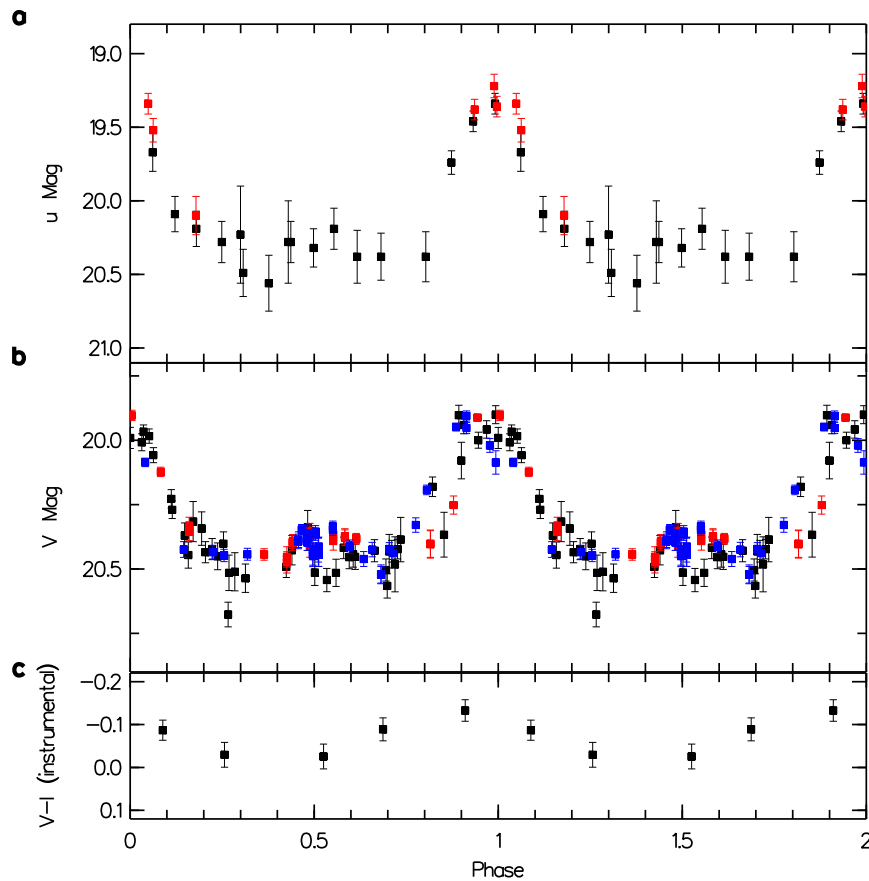


Extended Data Figure 6 | Orbital and superorbital periods. Best-fit orbital ($P_{\text{orb}} = 63.52$ d) and superorbital ($P_s = 2,620$ d ≈ 7.2 yr) solutions accounting for the periodic phase jitter of times of optical/ultraviolet photometric maxima. Error bars show 1σ statistical errors. Five times at which maximum light

occurred were extracted from the V-band photometry at MJD 53314.8 ± 3 , 53636.4 ± 3 , 53699.4 ± 2 (Las Campanas) and MJD 55532.8 ± 3 , 55788.0 ± 2 (ESO VLT) and four from the UVOT photometry at MJD 55468.0 ± 2 (u), 56175.0 ± 3 (u), 56243.0 ± 2 (u , $um2$, $uw2$), 56303.0 ± 2 (u , $um2$, $uw2$).

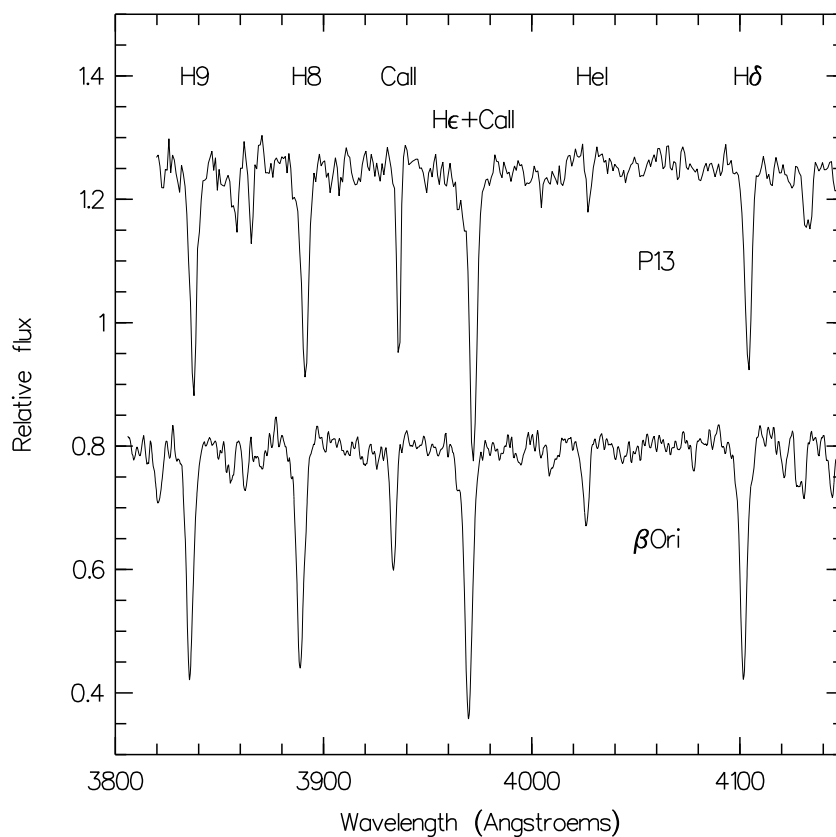


Extended Data Figure 7 | Folded radial velocity curves. He II $\lambda 4,686$ (a) and Balmer absorption (b) radial velocity curves folded with the best-fit combination of orbital and superorbital periods ($P_{\text{orb}} = 63.52$ d and $P_s = 2,620$ d). Phase 0 corresponds to maximum light. Blue, ESO 2009; black, ESO 2010; red, ESO 2011. The He II line displays a clear velocity change with orbital phase. Error bars in all panels show statistical uncertainties at the 1σ level.



Extended Data Figure 8 | Folded light curves. u , V and $V-I$ light curves folded with the best-fit combination of orbital and superorbital periods ($P_{\text{orb}} = 63.52$ d and $P_s = 2,620$ d). Phase 0 corresponds to the predicted time of maximum optical light. Note the different scales used to plot the u and V light curves. **a**, u : black, 2010 X-ray bright-state run; red, data acquired during the

faint-X-ray state in the time interval 2012 September 2 to 2013 January 12. **b**, V : black, Las Campanas; red, ESO 2010 (X-ray bright); blue, ESO 2011 (X-ray faint). **c**, Binned $V-I$ light curve. The $V-I$ index is plotted on an arbitrary scale. Error bars in all panels show statistical uncertainties at the 1σ level.



Extended Data Figure 9 | High-order Balmer lines. Normalized mean P13 spectrum away from maximum light ($V \geq 20.3$ mag, shifted up by 0.25 for clarity) compared to that of the B8Ia star β Orionis (shifted down by 0.2). The equivalent widths of the high-order Balmer lines H8 and H9 are almost identical in the two stars. The higher interstellar absorption towards P13 than

towards Orion is responsible for the stronger Ca II line in the ULX spectrum. Residual Balmer emission already adds to the photospheric H δ absorption line and the He I line is blended with the Ca II interstellar line. The mean equivalent width of the H8 and H9 Balmer lines are 1.52 ± 0.09 and 1.56 ± 0.08 for β Ori and P13, respectively (1σ errors), consistent with no line veiling in P13.

Extended Data Table 1 | X-ray spectral fits

Parameter	Value in C03	Value in X13
Power law ($E \geq 1.5$ keV)		
$n_{\text{H,int}}$ (10^{20} cm^{-2})	[0.0]	[0.0]
Γ	$1.07^{+0.06}_{-0.06}$	$1.24^{+0.03}_{-0.03}$
χ^2/dof	0.91 (167.7/185)	1.24 (458.7/367)
P(H0)	0.81	7.8×10^{-4}
Broken power law ($E \geq 1.5$ keV)		
$n_{\text{H,int}}$ (10^{20} cm^{-2})	[0.0]	[0.0]
Γ_{soft}	$0.97^{+0.08}_{-0.11}$	$1.19^{+0.05}_{-0.05}$
Γ_{hard}	$1.50^{+0.64}_{-0.30}$	$2.60^{+0.37}_{-0.29}$
E_{br} (keV)	$4.2^{+1.2}_{-1.3}$	$5.45^{+0.39}_{-0.31}$
χ^2/dof	0.87 (158.9/183)	0.99 (361.6/365)
P(H0)	0.90	0.54
Power law		
$n_{\text{H,int}}$ (10^{20} cm^{-2})	$6.0^{+1.4}_{-1.3}$	$4.8^{+0.8}_{-0.7}$
Γ	$1.18^{+0.05}_{-0.05}$	$1.45^{+0.03}_{-0.03}$
χ^2/dof	1.03 (268.9/261)	1.25 (751.3/603)
P(H0)	0.35	3.4×10^{-5}
Broken power law		
$n_{\text{H,int}}$ (10^{20} cm^{-2})	$5.8^{+1.5}_{-1.4}$	$3.4^{+0.8}_{-0.8}$
Γ_{soft}	$1.17^{+0.06}_{-0.06}$	$1.37^{+0.04}_{-0.04}$
Γ_{hard}	1.52^{+*}_{-*}	$2.75^{+0.59}_{-0.44}$
E_{br} (keV)	$5.1^{+*}_{-1.6}$	$6.16^{+0.52}_{-0.44}$
χ^2/dof	1.03 (267.4/259)	1.16 (698.7/601)
P(H0)	0.35	3.5×10^{-3}

Parameter	Value in C03	Value in X13
Disk blackbody + broken power law		
$n_{\text{H,int}}$ (10^{20} cm^{-2})	$7.6^{+3.6}_{-3.1}$	$7.8^{+2.6}_{-2.2}$
kT_{in} (keV)	$0.40^{+0.22}_{-0.14}$	$0.29^{+0.05}_{-0.04}$
N_{dbb}	$0.58^{+1.84}_{-0.44}$	$1.64^{+2.02}_{-0.86}$
Γ_{soft}	$0.71^{+0.29}_{-0.55}$	$1.08^{+0.09}_{-0.11}$
Γ_{hard}	$1.45^{+0.33}_{-0.23}$	$2.62^{+0.33}_{-0.29}$
E_{br} (keV)	$3.8^{+1.1}_{-0.7}$	$5.38^{+0.33}_{-0.38}$
χ^2/dof	0.98 (250.9/257)	1.03 (618.5/599)
P(H0)	0.60	0.28
Comptonization spectrum <i>diskbb+comptt</i>		
$n_{\text{H,int}}$ (10^{20} cm^{-2})	$0.0^{+11.0}_{-0.0}$	$6.4^{+1.6}_{-1.6}$
$kT_{\text{in}} \equiv kT_0$ (keV)	$0.20^{+0.08}_{-0.05}$	$0.32^{+0.03}_{-0.07}$
N_{dbb}	$0.0^{+5.9}_{-0.0}$	$1.6^{+0.8}_{-0.5}$
kT_e (keV)	$2.11^{+0.37}_{-0.23}$	$1.67^{+0.20}_{-0.12}$
τ	$11.7^{+1.2}_{-1.2}$	$14.6^{+0.7}_{-2.6}$
χ^2/dof	0.96 (247.1/258)	1.03 (619.1/600)
P(H0)	0.68	0.29
$L_{0.3-10\text{keV}}$ ($10^{39} \text{ erg s}^{-1}$)	$3.5^{+0.1}_{-0.1}$	$2.0^{+0.1}_{-0.1}$
Irradiated Comptonized disk <i>diskir</i>		
$n_{\text{H,int}}$ (10^{20} cm^{-2})	$2.2^{+2.3}_{-1.6}$	$6.5^{+3.2}_{-2.4}$
kT_{in} (keV)	$0.08^{+0.09}_{-0.02}$	$0.16^{+0.06}_{-0.03}$
Γ	$1.26^{+0.02}_{-0.03}$	$1.39^{+0.04}_{-0.09}$
kT_e (keV)	$1.64^{+0.12}_{-0.11}$	$1.80^{+0.12}_{-0.16}$
L_c/L_d	324^{+211}_{-163}	$7.4^{+20}_{-3.4}$
N_{dbb}	$7.5^{+4.8}_{-2.6}$	$11.9^{+8.0}_{-6.7}$
χ^2/dof	0.97 (249.3/258)	1.03 (617.8/600)
P(H0)	0.64	0.30
$L_{0.3-10\text{keV}}$ ($10^{39} \text{ erg s}^{-1}$)	$3.4^{+0.1}_{-0.1}$	$2.0^{+0.1}_{-0.1}$

Spectral fits to Chandra (C03) and XMM-Newton (X13) bright-state X-ray spectra, from 2003 and 2013, respectively. All errors are given at the 90% confidence level. $P(H0)$ is the null hypothesis probability. In all cases, we assume a fixed Galactic column density $n_{\text{H}} = 1.2 \times 10^{20} \text{ cm}^{-2}$ (tbabs model) in addition to a fitted intrinsic (NGC 7793 + local) absorption column. For the diskir model, we assumed³⁰ $f_{\text{in}} = 0.1$, $r_{\text{irr}} = 1.2$, $f_{\text{out}} = 0.005$ and $\log(r_{\text{out}}) = 5.0$.

30. Gierliński, M., Done, C. & Page, K. Reprocessing of X-rays in the outer accretion disc of the black hole binary XTE J1817–330. *Mon. Not. R. Astron. Soc.* **392**, 1106–1114 (2009).

An ultraluminous X-ray source powered by an accreting neutron star

M. Bachetti^{1,2}, F. A. Harrison³, D. J. Walton³, B. W. Grefenstette³, D. Chakrabarty⁴, F. Fürst³, D. Barret^{1,2}, A. Beloborodov⁵, S. E. Boggs⁶, F. E. Christensen⁷, W. W. Craig⁸, A. C. Fabian⁹, C. J. Hailey¹⁰, A. Hornschemeier¹¹, V. Kaspi¹², S. R. Kulkarni³, T. Maccarone¹³, J. M. Miller¹⁴, V. Rana³, D. Stern¹⁵, S. P. Tendulkar³, J. Tomsick⁶, N. A. Webb^{1,2} & W. W. Zhang¹¹

The majority of ultraluminous X-ray sources are point sources that are spatially offset from the nuclei of nearby galaxies and whose X-ray luminosities exceed the theoretical maximum for spherical infall (the Eddington limit) onto stellar-mass black holes^{1,2}. Their X-ray luminosities in the 0.5–10 kiloelectronvolt energy band range from 10^{39} to 10^{41} ergs per second³. Because higher masses imply less extreme ratios of the luminosity to the isotropic Eddington limit, theoretical models have focused on black hole rather than neutron star systems^{1,2}. The most challenging sources to explain are those at the luminous end of the range (more than 10^{40} ergs per second), which require black hole masses of 50–100 times the solar value or significant departures from the standard thin disk accretion that powers bright Galactic X-ray binaries, or both. Here we report broadband X-ray observations of the nuclear region of the galaxy M82 that reveal pulsations with an average period of 1.37 seconds and a 2.5-day sinusoidal modulation. The pulsations result from the rotation of a magnetized neutron star, and the modulation arises from its binary orbit. The pulsed flux alone corresponds to an X-ray luminosity in the 3–30 kiloelectronvolt range of 4.9×10^{39} ergs per second. The pulsating source is spatially coincident with a variable source⁴ that can reach an X-ray luminosity in the 0.3–10 kiloelectronvolt range of 1.8×10^{40} ergs per second¹. This association implies a luminosity of about 100 times the Eddington limit for a 1.4-solar-mass object, or more than ten times brighter than any known accreting pulsar. This implies that neutron stars may not be rare in the ultraluminous X-ray population, and it challenges physical models for the accretion of matter onto magnetized compact objects.

The brightest accretion-powered X-ray pulsars, A0538-66⁵, SMC X-1⁶ and GRO J1744-28⁷, are variable, with reported X-ray luminosities up to $L_X(2-20 \text{ keV}) \approx 10^{39} \text{ erg s}^{-1}$, which are at the low end of the range that defines ultraluminous X-ray sources (ULXs). Such luminosities, exceeding the Eddington limit for a neutron star by a factor of about six, can be understood⁸ as resulting from accretion of material at moderately super-Eddington rates through a disk that couples to the neutron star's strong dipolar magnetic field (surface fields of $B \approx 10^{12} \text{ G}$). At the magnetospheric (Alfvén) radius, material is funnelled along the magnetic axis, radiation escapes from the column's side, and radiation pressure is ineffective at arresting mass transfer^{9,10}. Explaining ULXs that have $L_X > 10^{39} \text{ erg s}^{-1}$ with a $\sim 1 M_\odot$ compact object using typical accreting pulsar models is however extremely challenging⁸.

The NuSTAR (Nuclear Spectroscopic Telescope Array) high energy X-ray mission¹¹ observed the galaxy M82 (at distance $d \approx 3.6 \text{ Mpc}$) seven times between 2014 January 23 and 2014 March 06 as part of a follow-up campaign of the supernova SN 2014J. The galaxy's disk contains several

ULXs, the most luminous being M82 X-1¹², which can reach $L_X(0.3-10 \text{ keV}) \approx 10^{41} \text{ erg s}^{-1}$, and the second brightest being a transient, M82 X-2 (also referred to as X42.3+59¹³), which has been observed to reach^{4,14} $L_X(0.3-10 \text{ keV}) \approx 1.8 \times 10^{40} \text{ erg s}^{-1}$. The two sources are separated by 5", and so can only be clearly resolved by the Chandra X-ray telescope. During the M82 monitoring campaign, NuSTAR observed bright emission from the nuclear region containing the two ULXs. The region shows moderate flux variability at the 20% level during the first 22 days of observation. The flux was then found to have decreased by 60% by the time of the final observation ~ 20 days later. The peak X-ray flux, $F_X(3-30 \text{ keV}) = (2.33 \pm 0.01) \times 10^{-11} \text{ erg cm}^{-2} \text{ s}^{-1}$ (90% confidence; Fig. 1) corresponds to a total 3–30 keV luminosity assuming isotropic emission of $3.7^{+0.01}_{-0.02} \times 10^{40} \text{ erg s}^{-1}$.

A timing analysis revealed a narrow peak just above the noise at a frequency of $\sim 0.7 \text{ Hz}$ in the power density spectrum. An accelerated epoch folding search¹⁵ on overlapping 30-ks intervals of data found coherent pulsations with a mean period P of 1.37 s modulated with a sinusoidal period of 2.53 days throughout the 10-day interval starting at modified Julian day (MJD) 56691 (2014 February 03), and also in the last observation at MJD 56720 (Fig. 1). The statistical significance of the pulsations is $\sim 13\sigma$ during the most significant 30-ks segments, and $> 30\sigma$ for the entire observation. A refined analysis (see Methods) subsequently enabled the detection of pulsations over a longer interval beginning on MJD 56686. The pulsed flux is variable, ranging from 5% to 13% in the 3–30 keV range, and from 8% to 23% in the 10–30 keV range; see Fig. 1c. Although the pulsed flux increases with energy, this may result from a reduction in the contamination from other sources in the point-spread function rather than from a true increase in pulsed fraction. The maximum pulsed luminosity of the periodic source, which we will hereafter refer to as NuSTAR J095551+6940.8, is $4.9^{+0.02}_{-0.03} \times 10^{39} \text{ erg s}^{-1}$ (3–30 keV).

Analysis of the period modulation yields a near-circular orbit (upper eccentricity limit of 0.003; see Methods) with a projected semi-major axis of 22.225(4) light s (1σ error). In addition to the orbital modulation, a linear spin-up of the pulsar is evident, with a period derivative $\dot{P} \approx -2 \times 10^{-10} \text{ s s}^{-1}$ over the interval from MJD 56696 to 56701 when the pulse detection is most significant. Phase-connecting the observations enables detection of a changing spin-up rate over a longer timespan (Fig. 2a) as well as erratic variations likely to be related to local changes in the torque on the neutron star applied by accreting matter¹⁶.

Chandra observed M82 from MJD 56690.8396 to 56692.618, during an epoch when pulsations are detected. Only two sources in the Chandra image, M82 X-1 and M82 X-2, are sufficiently bright to be the counterpart of NuSTAR J095551+6940.8 (see Methods). We find the centroid of the pulsed emission (right ascension $\alpha = 09 \text{ h } 55 \text{ min } 51.05 \text{ s}$, declination

¹Université de Toulouse, UPS-OMP, Institut de Recherche en Astrophysique et Planétologie, 9, Avenue du Colonel Roche, BP 44346, 31028 Toulouse Cedex 4, France. ²CNRS, Institut de Recherche en Astrophysique et Planétologie, 9, Avenue du Colonel Roche, BP 44346, 31028 Toulouse Cedex 4, France. ³Cahill Center for Astrophysics, 1216 East California Boulevard, California Institute of Technology, Pasadena, California 91125, USA. ⁴MIT Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ⁵Physics Department, Columbia University, 538 West 120th Street, New York, New York 10027, USA. ⁶Space Sciences Laboratory, University of California, Berkeley, California 94720, USA. ⁷DTU Space, National Space Institute, Technical University of Denmark, Elektrovej 327, DK-2800 Lyngby, Denmark. ⁸Lawrence Livermore National Laboratory, Livermore, California 94550, USA. ⁹Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK. ¹⁰Columbia Astrophysics Laboratory, Columbia University, New York, New York 10027, USA. ¹¹NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. ¹²Department of Physics, McGill University, Montreal, Quebec H3A 2T8, Canada. ¹³Department of Physics, Texas Tech University, Lubbock, Texas 79409, USA. ¹⁴Department of Astronomy, University of Michigan, 500 Church Street, Ann Arbor, Michigan 48109-1042, USA. ¹⁵Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA.

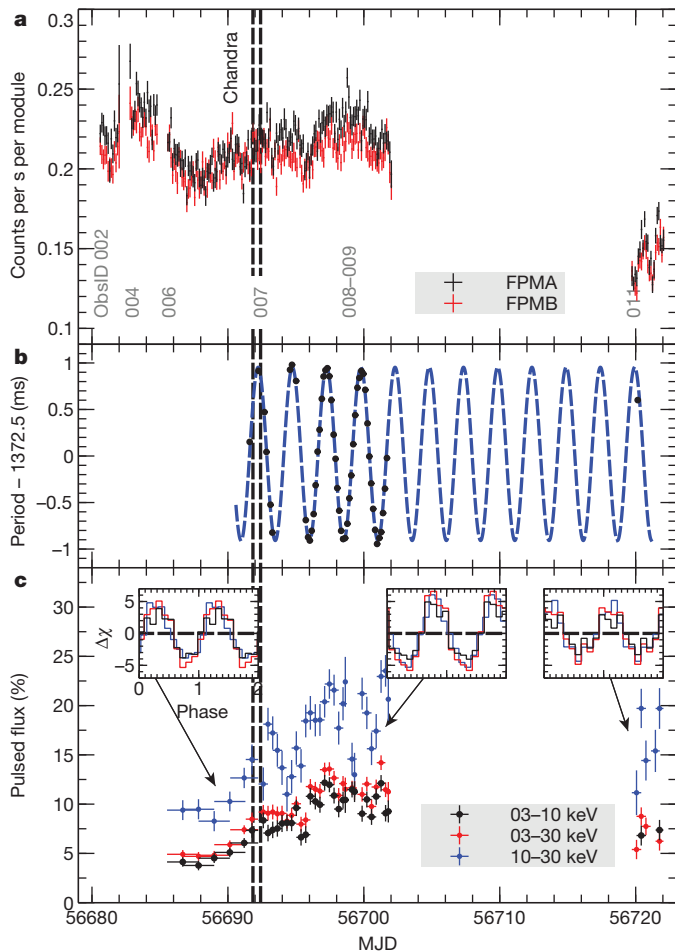


Figure 1 | The X-ray light curve and pulsations from the region containing NuSTAR J095551+6940.8. **a**, The background-subtracted 3–30 keV light curve extracted from a 70''-radius region around the position of NuSTAR J095551+6940.8. Black and red indicate the count rate from each of the two NuSTAR focal plane modules (FPMA and FPMB; 1 σ errors). The vertical grey labels indicate different observations. **b**, Detection of the pulse period. Data (black points) are fitted using the best sinusoidal ephemeris (blue dashed line). The mean period is 1.37252266(12) seconds, with an orbital modulation period of 2.51784(6) days. The dashed vertical lines through all panels delineate the contemporaneous Chandra observation. **c**, Pulsed flux as a fraction of the emission from the 70'' region. Insets, pulse profile at indicated points, normalized so that $\sigma = 1$.

$\delta = +69^\circ 40' 47.9''$) to be consistent with the location of M82 X-2 (Fig. 3). Monitoring by the Swift satellite establishes that the decrease in the nuclear region flux seen during observation (ObsID) 011 (see Extended Data Table 1) is due to fading of M82 X-1. The persistence of pulsations during this time further secures the association of the pulsating source, NuSTAR J095551+6940.8, with M82 X-2. We derive a flux $F_X(0.5\text{--}10\text{ keV}) = 4.07 \times 10^{-12} \text{ erg cm}^{-2} \text{ s}^{-1}$, and an unabsorbed luminosity of $L_X(0.5\text{--}10\text{ keV}) = (6.6 \pm 0.1) \times 10^{39} \text{ erg s}^{-1}$ for M82 X-2 during the Chandra observation.

The detection of coherent pulsations, a binary orbit, and spin-up behaviour indicative of an accretion torque unambiguously identify NuSTAR J095551+6940.8 as a magnetized neutron star accreting from a stellar companion. The highly circular orbit suggests the action of strong tidal torques, which, combined with the high luminosity, point to accretion via Roche lobe overflow. The orbital parameters give a Newtonian mass function $f = 2.1 M_\odot$ (here M_\odot indicates the solar mass), and the lack of eclipses and assumption of a Roche-lobe-filling companion constrain the inclination to be $i < 60^\circ$. The corresponding minimum companion mass assuming a $1.4 M_\odot$ neutron star is $M_c > 5.2 M_\odot$, with radius $R_c > 7 R_\odot$.

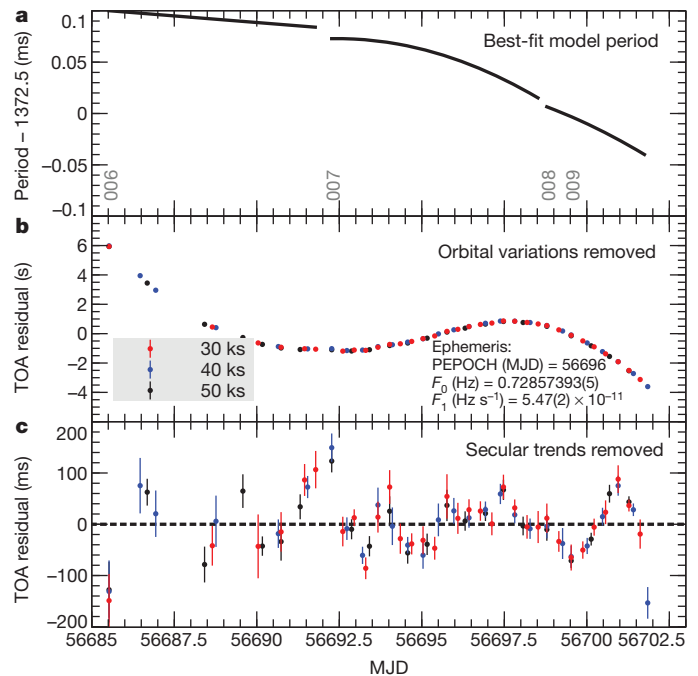


Figure 2 | The spin-up behaviour of NuSTAR J095551+6940.8. **a**, The residual period after correcting for the sinusoidal orbital modulation given in Extended Data Table 2. The period, displayed through the best-fit in Extended Data Table 3, decreases consistently, but the spin-up rate is changing. **b**, Time of arrival (TOA) residuals after removing the best-fit sinusoidal orbital modulation and a constant period derivative (the parameters are shown in commonly used units²⁶). PEPOCH, F_0 and F_1 are the reference time and the pulse frequency and its derivative, respectively. There is a clear trend independent of the choice of time binning (30, 40 or 50 ks) that results from the variable spin-up. **c**, Residuals after a smooth curve is fitted to the TOA residuals. Residual noise remains in the TOAs at the 100 ms level (1 σ uncertainties).

It is challenging to explain the high luminosity using standard models for accreting magnetic neutron stars. Adding the Chandra-measured $E < 10\text{ keV}$ luminosity to the $E > 10\text{ keV}$ pulsed flux (NuSTAR cannot directly spatially resolve the ULX), NuSTAR J095551+6940.8 has a luminosity $L_X(0.5\text{--}30\text{ keV}) \approx 10^{40} \text{ erg s}^{-1}$. Theoretically, the X-ray luminosity depends strongly on the magnetic field and the geometry of the accretion channel, being largest for a thin, hollow funnel that can result from the coupling of a disk onto the magnetic field¹⁰. A limiting luminosity $L_X \approx \frac{l_0}{2\pi d_0} L_{\text{Edd}}$, where l_0 is the arc length of the funnel, d_0 its thickness, and L_{Edd} the Eddington luminosity, can be reached if the magnetic field is high enough ($B \geq 10^{13} \text{ G}$) to contain the accreting gas column⁸. Ratios of $l_0/d_0 \approx 40$ are plausible, so that the limiting luminosity can reach $L_X \approx 10^{39} \text{ erg s}^{-1}$, implying mass transfer rates exceeding the Eddington value by many times. Beyond this, additional factors increasing L_X could result from increased L_{Edd} due to very high ($B > 10^{14} \text{ G}$) fields, which can reduce the electron scattering opacity¹⁷, and/or a heavy neutron star. Some geometric beaming is also likely to be present.

This scenario is, however, difficult to reconcile with the measured rate of spin-up. The spin-up results from the torque applied by accreting material threading onto the magnetic field^{18,19}. NuSTAR J095551+6940.8 is likely to be in spin equilibrium, given the short spin-up timescale, $P/\dot{P} \approx 300 \text{ yr}$. Near equilibrium, the magnetosphere radius, r_m , is comparable to the co-rotation radius (the radius where a Keplerian orbit co-rotates with the neutron star):

$$r_{\text{co}} = \left(\frac{GM_{\text{NS}} P^2}{4\pi^2} \right)^{1/3} = 2.1 \times 10^8 \left(\frac{M_{\text{NS}}}{1.4 M_\odot} \right)$$

Here G is the gravitational constant, M_{NS} is the neutron star mass, and r_{co} is in cm. With this assumption we can convert the measured torque,

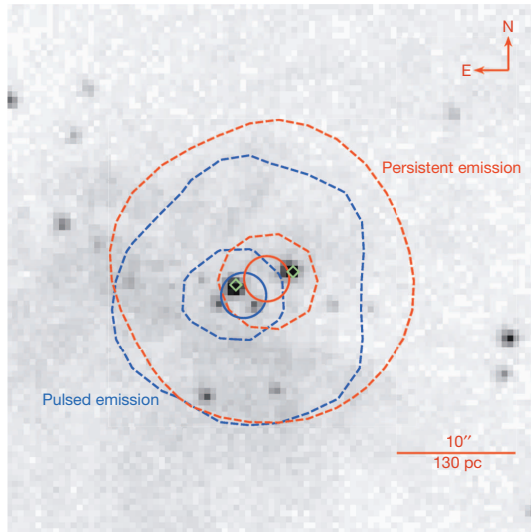


Figure 3 | The counterpart of NuSTAR J095551+6940.8. The greyscale image shows a $45'' \times 45''$ Chandra image of the galaxy's centre. Green diamonds mark the locations of M82 X-1 and X-2. NuSTAR 10–40 keV intensity contours (dashed) (50% and 90% levels) are shown for the pulsed (blue) and the persistent (red) emission. Solid error circles indicate the 3σ statistical uncertainty on the centroid locations (see Methods). The pulsed emission centroid is consistent with the location of M82 X-2, and the centroid of the persistent emission is between M82 X-1 and X-2, indicating that there is additional persistent emission from X-2 as well as the persistent emission from X-1.

$\tau = 2\pi\dot{I}\dot{\nu} = 6 \times 10^{35} I_{45} \text{ g cm}^2 \text{ s}^{-2}$ (where I_{45} is the neutron star moment of inertia I (in units of 10^{45} g cm^2) and $\dot{\nu}$ is the measured frequency derivative) into a rate of matter magnetically channelled onto the pulsar:

$$\dot{M}_{\text{mag}} = 5 \times 10^{-8} M_{1.4}^{-\frac{1}{2}} \left(\frac{r_m}{r_{\text{co}}} \right)^{-\frac{1}{2}} M_{\odot} \text{ yr}^{-1} = 2.5 \left(\frac{r_m}{r_{\text{co}}} \right)^{-\frac{1}{2}} \dot{M}_{\text{Edd}}$$

where $M_{1.4}$ is the mass in units of 1.4 solar masses and \dot{M}_{Edd} is the Eddington accretion rate. From the spin-up we therefore find an accretion rate that is only a few times higher than Eddington, independent of any assumption about the pulsar magnetic field. For an equilibrium spin period of 1.37 s and $\dot{M} \approx \dot{M}_{\text{Edd}}$, the implied magnetic field is $B \approx 10^{12} \text{ G}$, typical of accreting pulsars in high mass X-ray binaries, and too low to have an appreciable effect on L_{Edd} . It is possible that the current \dot{M} is significantly larger than the average for this system, so that $r_m < r_{\text{co}}$, increasing \dot{M}_{mag} . For $\dot{M} = 100\dot{M}_{\text{avg}}$, \dot{M}_{mag} is increased tenfold, so that only moderate geometric beaming is required to explain the observed luminosity. A fan beam geometry⁹ viewed at a favourable angle could in this case produce the observed pulse profile (Fig. 1) and provide the requisite moderate collimation.

The discovery of an ultraluminous pulsar has implications for understanding the ULX population. The fraction of ULXs powered by neutron stars must now be considered highly uncertain. M82 X-2 has been extensively studied^{4,14}, but pulsations have until now eluded detection owing to the limited timing capabilities of sensitive X-ray instruments, the transient nature of the pulsations, and the large amplitude of the orbital motion. Pulsars may indeed not be rare among the ULX population.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 June; accepted 6 August 2014.

1. Roberts, T. P. X-ray observations of ultraluminous X-ray sources. *Astrophys. Space Sci.* **311**, 203–212 (2007).
2. Liu, J.-F., Bregman, J. N., Bai, Y., Justham, S. & Crowther, P. Puzzling accretion onto a black hole in the ultraluminous X-ray source M 101 ULX-1. *Nature* **503**, 500–503 (2013).
3. Feng, H. & Soria, R. Ultraluminous X-ray sources in the Chandra and XMM-Newton era. *New Astron. Rev.* **55**, 166–183 (2011).
4. Feng, H., Rao, F. & Kaaret, P. Discovery of millihertz X-ray oscillations in a transient ultraluminous X-ray source in M82. *Astrophys. J.* **710**, L137–L141 (2010).
5. Skinner, G. K. *et al.* Discovery of 69 ms periodic X-ray pulsations in A0538–66. *Nature* **297**, 568–570 (1982).
6. Lucke, R., Yentis, D., Friedman, H., Fritz, G. & Shulman, S. Discovery of X-ray pulsations in SMC X-1. *Astrophys. J.* **206**, L25–L28 (1976).
7. Kouveliotou, C. *et al.* A new type of transient high-energy source in the direction of the Galactic Centre. *Nature* **379**, 799–801 (1996).
8. Basko, M. M. & Sunyaev, R. A. The limiting luminosity of accreting neutron stars with magnetic fields. *Mon. Not. R. Astron. Soc.* **175**, 395–417 (1976).
9. Gnedin, Y. N. & Sunyaev, R. A. The beaming of radiation from an accreting magnetic neutron star and the X-ray pulsars. *Astron. Astrophys.* **25**, 233–239 (1973).
10. Basko, M. M. & Sunyaev, R. A. Radiative transfer in a strong magnetic field and accreting X-ray pulsars. *Astron. Astrophys.* **42**, 311–321 (1975).
11. Harrison, F. A. *et al.* The Nuclear Spectroscopic Telescope Array (NuSTAR) high-energy X-ray mission. *Astrophys. J.* **770**, 103–122 (2013).
12. Kaaret, P. *et al.* Chandra High-Resolution Camera observations of the luminous X-ray source in the starburst galaxy M82. *Mon. Not. R. Astron. Soc.* **321**, L29–L32 (2001).
13. Kaaret, P., Simet, M. G. & Lang, C. C. A 62 day X-ray periodicity and an X-ray flare from the ultraluminous X-ray source in M82. *Astrophys. J.* **646**, 174–183 (2006).
14. Kong, A. K. H., Yang, Y. J., Hsieh, P. Y., Mak, D. S. Y. & Pun, C. S. J. The ultraluminous X-ray sources near the center of M82. *Astrophys. J.* **671**, 349–357 (2007).
15. Ransom, S. M. *New Search Techniques for Binary Pulsars*. PhD thesis, Harvard Univ. (2001).
16. Bildsten, L. *et al.* Observations of accreting pulsars. *Astrophys. J. Suppl. Ser.* **113**, 367–408 (1997).
17. Canuto, V., Lodenquai, J. & Ruderman, M. Thomson scattering in a strong magnetic field. *Phys. Rev. D* **3**, 2303–2308 (1971).
18. Pringle, J. E. & Rees, M. J. Accretion disc models for compact X-ray sources. *Astron. Astrophys.* **21**, 1–9 (1972).
19. Ghosh, P. & Lamb, F. K. Accretion by rotating magnetic neutron stars. III — Accretion torques and period changes in pulsating X-ray sources. *Astrophys. J.* **234**, 296–316 (1979).

Acknowledgements This work was supported by NASA (grant no. NNG08FD60C), and made use of data from the Nuclear Spectroscopic Telescope Array (NuSTAR) mission, a project led by Caltech, managed by the Jet Propulsion Laboratory and funded by NASA. We thank the NuSTAR operations, software and calibration teams for support with execution and analysis of these observations. This work made use of data supplied by the UK Swift Science Data Centre at the University of Leicester. M.B. thanks the Centre National d'Études Spatiales (CNES) and the Centre National de la Recherche Scientifique (CNRS) for support. Line plots were done using Veusz software by J. Sanders.

Author Contributions M.B., reduction and timing analysis of the NuSTAR observations, interpretation of results, manuscript preparation; F.A.H., interpretation of results, manuscript preparation; D.J.W., NuSTAR and Chandra spectroscopy, point source analysis; B.W.G., NuSTAR image analysis; D.C., accretion torque analysis, interpretation; F.F., verification of timing analysis, interpretation; D.B., A.B., A.C.F., A.H., V.M.K., T.M., J.T., interpretation of results and manuscript review; S.B., F.C., W.W.C., C.J.H., D.S., S.P.T., N.W., W.W.Z., manuscript review.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.B. (mbachett@oa-cagliari.inaf.it) and F.A.H. (fiona@srll.caltech.edu).

METHODS

Observations and preliminary data reduction. NuSTAR¹¹ observed the M82 field 7 times between 2014 January 23 and 2014 March 06 (see Extended Data Table 1 for details), for a total exposure of 1.91 Ms. We used the NuSTAR data analysis software (NuSTARDAS) version 1.2.0 and NuSTAR CALDB version 20130509 with the standard filters to obtain good time intervals, excluding the periods where the source was occulted by the Earth or was transiting through the South Atlantic Anomaly. NuSTAR records event arrival times with a resolution of 10 μ s. Clock drifts, mostly due to temperature, are recorded at each ground station passage into a clock correction file that is updated monthly. The final time accuracy that can be reached by applying these clock corrections is ~ 2 ms. We applied these corrections to obtain Solar System barycentre corrected event times using the general-purpose FTOOL barycorr and the NuSTAR clock correction file (v030).

Chandra²⁰ observed the M82 field using the ACIS-S detector in timed exposure (TE)/VFAINT mode between 2014-02-13 20:09:32 and 2014-02-04 14:50:00 for a total exposure of 47 ks. We reduced the data with the standard pipeline in the Chandra Interactive Analysis of Observations software package (CIAO, version 4.6), filtering the data for periods of high background to produce a cleaned event list. We extracted point source spectra from circular regions of radius ~ 1 – $2''$, depending on the proximity of other nearby sources, while the background was estimated from a larger region of radius $\sim 35''$, free of contaminating point sources and away from the plane of the M82 galaxy. Spectra and instrumental responses were produced from the cleaned events with the CIAO SPECEXTRACT tool, and were corrected for the fraction of the PSF falling outside the source region.

The primary goal of the Chandra observation was to constrain any faint soft X-ray emission from the recent SN 2014J²¹, and so the observation was performed with the maximum frame-time of 3.2 s. Unfortunately, the two ULXs in M82 (X-1 and X-2) are sufficiently bright that the long frame integration time resulted in the Chandra spectra from these sources suffering from fairly severe pileup²² (pileup refers to the scenario in which more than one photon is incident on a detector pixel within one frame time, resulting in the spurious apparent detection of a single photon with the combined energy of the individual incident photons). The detected counts-per-frame is ~ 0.45 for both X-1 and X-2, which equates to a pileup fraction of ~ 20 – 30% based on the Chandra ABC guide to pileup (http://cxc.harvard.edu/ciao/download/doc/pileup_abc.pdf). This unfortunately prevents a straightforward estimation of the fluxes of these two sources from the Chandra data alone. We will return to this issue below. None of the other point sources are bright enough to suffer significantly from these effects.

There is a known flux calibration discrepancy of $\sim 10\%$ between NuSTAR and Chandra. We account for this discrepancy, and quote fluxes based on the NuSTAR zeropoint.

Swift monitored the field of SN 2014J regularly from 2014 January 23 through to the end of 2014 April. We reduced the data using the Swift mission automated data analysis server²³.

Timing analysis. To determine the mean period and orbital parameters we divided the observations into overlapping 30-ks intervals and ran an accelerated epoch folding search¹⁵ in each interval. To do this we used the software PRESTO¹⁵, originally designed for radio observations of pulsars. To adapt the NuSTAR data to this software we modified the makebininifscript by Abdo and Ray, originally written to use PRESTO with data from the Fermi satellite. To assess the quality of detection, accounting for possible difference in statistics between radio and X-ray data, we ran the same search with a random distribution of initial guesses for the period, far from the observed pulse period, and looked at the distribution of σ for the ‘best detections’. More than 99% of false detections were below a PRESTO value of 8σ , which we use to define an acceptable detection. With this criterion we detected pulsations over ~ 10 days of observations. We cross-checked the detection with independent software using ISIS²⁴.

The observed pulse period exhibits sinusoidal variations resulting from the binary orbital motion. Assuming a circular orbit, the observed period p_{obs} varies according to $p_{\text{obs}} = p_{\text{em}}[1 + X\Omega \sin[\Omega(t - T_{90})]]$, where p_{em} is the period in the system of reference co-moving with the pulsar, X is the projected semi-major axis of the orbit in light seconds, $\Omega = 2\pi/P_{\text{orb}}$ is the orbital angular velocity and T_{90} is the time of longitude 90° (or mid-eclipse) assuming the ascending node is at longitude 0.

During ObsIDs 008–009, the interval over which we fit for the orbital parameters, we find an additional trend in the phases which requires a period derivative $\dot{P} \approx -2 \times 10^{-10} \text{ s s}^{-1}$. NuSTAR’s relative timing accuracy is currently known to ~ 2 ms over time intervals comparable to these observations. There is evidence for clock drifts of ~ 0.4 ms on the ~ 97 min orbital timescale. Taking this drift into account, the observed period could have a systematic error of up to $\sim 10^{-7}$ s. This is orders of magnitude below the observed trend (~ 0.1 ms) that we associate with a period derivative.

The ephemeris determined above is not sufficiently precise to align the pulses throughout any of the observations. In order to refine it, we calculated times of arrival

(TOA) of the pulsations in each 30 ks interval and searched for an ephemeris that connects their phases²⁵. To do this, we used the software Tempo2²⁶. We found an orbital solution that aligns the TOAs to better than 36 ms (r.m.s.) from MJD 56696 to MJD 56701, the interval with the most significant detections.

We looked for signatures of eccentricity using the ELL1 model in Tempo2, adequate for low eccentricity orbits. We used TOAs calculated on different timescales (from 10 ks to 40 ks) in order to account for possible timescale-related effects. We measured values of the eccentricity that were always consistent with 0 within 2σ , with a maximum eccentricity of 0.002. We used this estimate to obtain a rough upper limit on the eccentricity of 0.003. The parameters of this orbital solution (determined for ObsIDs 008–009) can be found in Extended Data Table 2. Since the orbit is circular, we used the convention $T_0 = T_{\text{asc}}$ and so $T_{90} = T_{\text{asc}} + P_{\text{orb}}/4$, where T_0 , T_{asc} and T_{90} are the times of passage through periastron, ascending node and mean longitude 90° respectively.

We found that the above ephemeris was not valid before MJD 56696, with the residuals from the best fit rapidly departing from 0. Therefore, for the rest of the ObsIDs, we used PRESTO with the previously determined orbital solution, and searched the period– \dot{P} plane for more precise solutions inside individual ObsIDs. We determined values of period and period derivative that align the pulsation to better than 10% of the period inside single ObsIDs. By using each of these newly determined ephemerides as a first approximation and propagating to the earlier observation, we were able to detect the pulsation on a much longer interval, including the whole of ObsID 006. These local solutions are summarized in Extended Data Table 3, and are the used for the phase-resolved analysis below.

We looked for any modulation of the phases at the orbital frequency (or harmonics) that might hint at poorly determined orbital parameters or yet-unmodelled orbital parameters (for example, eccentricity, orbital period decay). We first subtracted the long-term trend of the phases with an *ad hoc* function, as implemented in Tempo2. The residuals of this curve have an erratic behaviour with quite large departures from a constant on timescales of several days. We calculated the Lomb-Scargle periodogram²⁷ of the residuals, looking for significant signatures at the orbital period or factors. We found no significant modulations of this kind. This timing noise is therefore not of orbital origin.

Counterpart identification. Chandra imaging: Extended Data Fig. 1 shows an image of the central region of M82 from Chandra in the 0.3–10 keV band made from the observation coincident with NuSTAR ObsID 006. Given the pileup issues suffered by X-1 and X-2, we focus here on estimating fluxes from the rest of the point source population. Spectra are extracted for each source individually following the method outlined above. Given that these sources are mostly rather faint, we systematically rebin their spectra to have a minimum of 5 counts per bin, and minimize the Cash statistic²⁸ when analysing these data. All spectral analysis is performed with XSPEC v12.8.1²⁹, and unless stated otherwise, quoted parameter uncertainties are the 90% confidence intervals for one parameter of interest.

We fit each source with two simple spectral models, the first an absorbed power law and the second an absorbed accretion disk (DISKBB³⁰ in XSPEC), focusing on the 1–10 keV bandpass to minimize the effects of the strong diffuse plasma emission from the M82 galaxy³¹, which peaks at very low energies. When applying the former model, we limit the photon index to be positive, and for the latter model, we limit the disk temperatures to be less than 5 keV. These simple models provide good phenomenological descriptions of the data for the majority of the sources, the only exception being source 15, and we compute fluxes from the 3–10 keV bandpass for each (that is, the common bandpass between Chandra and NuSTAR). For source 15, the Chandra data strongly require a second, soft X-ray component, and we additionally include a low temperature thermal plasma, modelled with the MEKAL code³², assuming solar abundances, and compute 3–10 keV fluxes based on these models. These Chandra fluxes are compared to the total NuSTAR flux in the 3–10 keV bandpass of $(1.50 \pm 0.03) \times 10^{-11} \text{ erg cm}^{-2} \text{ s}^{-1}$, which we compute with a phenomenological combination of a Gaussian iron K α emission line and a Comptonized continuum model (CompTT³³) fitted to the portion of the NuSTAR data simultaneous with the Chandra observation, extracted from the $70''$ region shown in Extended Data Fig. 1.

Extended Data Table 4 provides the results of the power-law fits for the brightest five sources in this population. The disk-blackbody fluxes are typically within 30% of the powerlaw results (the latter generally being higher). Even taking the conservative approach and adopting the slightly higher powerlaw fluxes, none of these sources contribute more than 5% of the 3–10 keV NuSTAR flux. Given the observed 3–10 keV pulse fraction at the time of the Chandra observation, it is clear that only X-1 and X-2 are bright enough to be the origin of these pulsations.

NuSTAR astrometry: we use the ‘persistent’ emission to determine an absolute astrometric correction, which we then apply to the ‘pulse-on’ images to determine the location of the pulsed emission. We assume that the persistent emission is composed equally of emission from M82 X-1 and X-2 since the contemporaneous Chandra imaging shows that these dominate the image and have comparable flux. To eliminate contamination from diffuse soft X-ray emission, we limit the image to the

energy range from 10 to 40 keV. We find that we can align the NuSTAR image with the mean position of M82 X-1 and X-2 by applying a shift in the image plane of two sky pixels (which have a plate scale of 2.45 arcsec per pixel), which is a reasonable shift given NuSTAR's astrometric accuracy¹¹.

After astrometric alignment we use the `gcntnd.pro` routine from the IDL `AstroLib` to determine the position of the pulsed emission using data within 30 arcsec of the peak brightness. This represents a net sample of 2,500 counts in the subtracted image.

To assess the statistical uncertainty on the source centroid, we perform a Monte Carlo simulation that distributes 2,500 counts in a simulated image according to the on-axis NuSTAR PSF. For each Monte Carlo run we compute the centroid as above and compare the result with the (known) input source location. The X and Y offset distributions from the Monte Carlo have means of 0 and 1σ widths of 0.235 pixels (0.577 arcsec), corresponding to a 90% confidence interval of ± 0.95 arcsec in both X and Y .

The resulting right ascension (α)/declination (δ) centroid of the NuSTAR pulsed image is $\alpha = 09^{\text{h}} 55^{\text{m}} 51.05^{\text{s}}$, $\delta = +69^{\circ} 40' 47.9''$, J2000, with a combined statistical and systematic error of $\pm 5''$ in the projected image plane.

Swift imaging: we perform imaging analysis on Swift observations made during ObsIDs 007 and 009, when pulsations are detected and the flux from the region is high (see Fig. 1). We use the Swift-XRT automated data processing pipeline²³, obtaining all observations from UT 2014 February 02 through 2014 February 11 and producing an image in the 5–10 keV band (Extended Data Fig. 2, left). We then compare this to the Swift snapshot observations extracted from 2014 March 07 through 2014 March 11 that occurred just after the ObsID 011, when the flux from the NuSTAR extraction region containing M82 X-1 and X-2 had decreased by 40%, but pulsations were still detected. When the flux is high the Swift imaging shows significant contributions from both M82 X-1 and X-2, but after ObsID 011 M82 X-2 is dominant in the Swift images. This is consistent with a significant decrease in flux of M82 X-1. The continued detection of pulsations indicates that they originate from M82 X-2.

Flux measurements and spectroscopy of M82 X-1 and M82 X-2. We estimate the 3–10 keV flux from M82 X-2 during our observations from the Chandra observation. As noted previously, the Chandra data for X-2 are quite piled-up, which results in strong degeneracies in the flux inferred from these data alone. By including the NuSTAR data, which does not suffer from these effects, it is possible to break this degeneracy. Since the NuSTAR aperture is much larger than the aperture used for the Chandra data, it includes the emission from a variety of other point sources, including X-1 (for which the Chandra data are also piled-up, resulting in similar issues), and also much of the diffuse M82 plasma emission. The contribution of these additional sources of emission to the NuSTAR data need to be accounted for before the flux of X-2 can reliably be estimated from the combined data set. Throughout all our analysis, we include a neutral absorption component fixed at the Galactic column in the direction of M82²⁴ ($N_{\text{H}} = 5.04 \times 10^{20} \text{ cm}^{-2}$) and additionally allow for variable neutral absorption intrinsic to M82.

Diffuse emission: we assess the contribution from the diffuse plasma emission, and extract an integrated Chandra spectrum from the 70" NuSTAR aperture with SPEXTRACT, excluding all the point sources identified in Extended Data Fig. 1. For the subsequent analysis, we follow ref. 31. We leave a detailed analysis of the diffuse emission during this observation for future studies; for our purposes, it is only important that a phenomenological model with three hot diffuse gas (MEKAL Xspec model) components provides a good description of the data. We find that this diffuse emission contributes $(20 \pm 1)\%$ of the observed 3–10 keV NuSTAR flux.

Other point sources: we model the emission from the additional X-ray source population (that is, excluding X-1 and X-2) for inclusion in the joint fitting. None of these sources are bright enough to significantly suffer from pileup. We first focus on the brightest two sources from this additional X-ray population, sources 15 and 18, and fit their full 0.3–10.0 keV Chandra spectra with a phenomenological model composed of a low-temperature MEKAL plasma (to account for the diffuse emission) and an accretion disk continuum.

For the rest of the fainter point source population, we combine their lower quality individual Chandra spectra into a single average spectrum, and simply model this to determine their contribution. We apply the same accretion disk continuum model

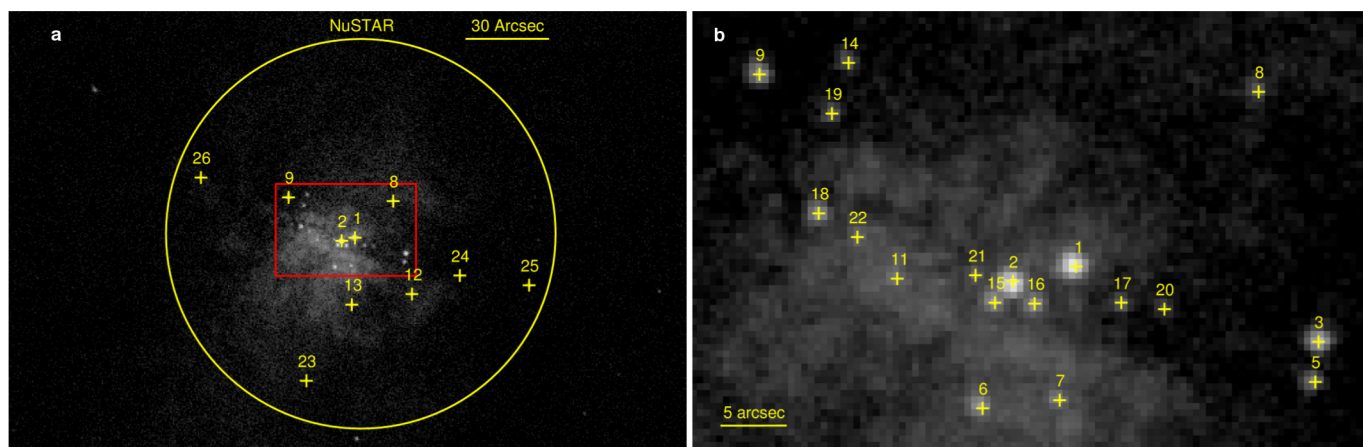
used above. The total contribution from point sources other than X-1 and X-2 is found to be $(17 \pm 1)\%$ of the observed 3–10 keV NuSTAR flux.

M82 X-1 and X-2: having characterized all the other sources of emission in the NuSTAR aperture, we can now constrain the fluxes of X-1 and X-2 (given the pileup issues in the Chandra data, both sources need to be analysed in this fashion simultaneously). We construct a 'contamination' model for the NuSTAR data obtained simultaneously with the Chandra observation, which is the sum of our characterizations of the diffuse emission and the other point sources, described above. The parameters of this contamination model are all set to the best-fit values obtained in our separate analysis of these various emission components, and are not allowed to vary.

To this model, we add two additional continuum components to represent X-1 and X-2, both absorbed. For the continuum, we use the CompTT model, as with different parameter combinations this model has the flexibility to provide both powerlaw-like and blackbody-like spectra, as required by the data. We then construct the same CompTT models for the Chandra data for X-1 and X-2 individually, and link the parameters for these components between the NuSTAR and the respective Chandra spectra. Finally, the Chandra models are modified by a pileup kernel²², which accounts for the spectral distortions introduced by these effects. The influence of these effects, characterized by the grade migration parameter α (ref. 22), is not known *a priori*, and is free to vary independently for X-1 and X-2. Although this pileup kernel can introduce strong degeneracies in the intrinsic flux inferred, this simultaneous modelling of the NuSTAR and the Chandra data, and the careful modelling of all the other sources of emission in the NuSTAR aperture, allows us to account for the pileup in the Chandra data, while also requiring that the total flux does not exceed that in the NuSTAR aperture, breaking this flux degeneracy.

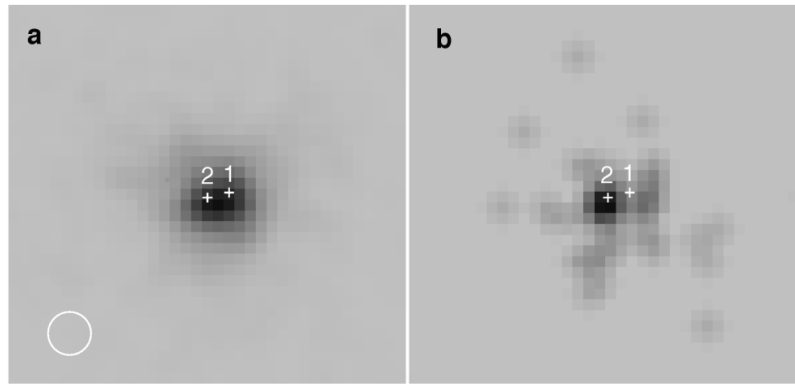
With this approach, we find that the 3–10 keV fluxes of X-1 and X-2 are relatively similar during this Chandra observation, contributing $42^{+3}_{-4}\%$ and $21^{+1}_{-1}\%$ of the 3–10 keV NuSTAR flux, corresponding to 3–10 keV luminosities of $(10 \pm 1) \times 10^{39} \text{ erg s}^{-1}$ and $4.9^{+0.6}_{-0.3} \times 10^{39} \text{ erg s}^{-1}$ respectively (for a distance to M82 of 3.6 Mpc). Using the measured spectrum and absorption column, we find an unabsorbed 0.5–10 keV luminosity of $6.6 \times 10^{39} \text{ erg s}^{-1}$ for X-2.

20. Weisskopf, M. C. *et al.* An overview of the performance and scientific results from the Chandra X-ray observatory. *Publ. Astron. Soc. Pacif.* **114**, 1–24 (2002).
21. Goobar, A. *et al.* The rise of SN 2014J in the nearby galaxy M82. *Astrophys. J.* **784**, L12 (2014).
22. Davis, J. E. Event pileup in charge-coupled devices. *Astrophys. J.* **562**, 575–582 (2001).
23. Evans, P. A. *et al.* Methods and results of an automatic analysis of a complete sample of Swift-XRT observations of GRBs. *Mon. Not. R. Astron. Soc.* **397**, 1177–1201 (2009).
24. Houck, J. C. & Denicola, L. A. ISIS: an Interactive Spectral Interpretation System for high resolution X-ray spectroscopy. *Astron. Data Analysis Softw. Syst.* **216**, 591–594 (2000).
25. Blandford, R. & Teukolsky, S. A. Arrival-time analysis for a pulsar in a binary system. *Astrophys. J.* **205**, 580–591 (1976).
26. Edwards, R. T., Hobbs, G. B. & Manchester, R. N. TEMPO2, a new pulsar timing package — II. The timing model and precision estimates. *Mon. Not. R. Astron. Soc.* **372**, 1549–1574 (2006).
27. Scargle, J. D. Studies in astronomical time series analysis. II — Statistical aspects of spectral analysis of unevenly spaced data. *Astrophys. J.* **263**, 835–853 (1982).
28. Cash, W. Parameter estimation in astronomy through application of the likelihood ratio. *Astrophys. J.* **228**, 939–947 (1979).
29. Arnaud, K. A. XSPEC: the first ten years. *Astron. Data Analysis Softw. Syst.* **101**, 17–20 (1996).
30. Mitsuda, K. *et al.* Energy spectra of low-mass binary X-ray sources observed from TENMA. *Astron. Soc. Jpn* **36**, 741–759 (1984).
31. Ranalli, P., Comastri, A., Origlia, L. & Maiolino, R. A deep X-ray observation of M82 with XMM-Newton. *Mon. Not. R. Astron. Soc.* **386**, 1464–1480 (2008).
32. Mewe, R., Gronenschild, E. H. B. M. & van den Oord, G. H. J. Calculated X-radiation from optically thin plasmas. V. *Astron. Astrophys. Suppl. Ser.* **62**, 197–254 (1985).
33. Titarchuk, L. Generalized Comptonization models and application to the recent high-energy observations. *Astrophys. J.* **434**, 570–586 (1994).
34. Kalberla, P. M. W. *et al.* The Leiden/Argentine/Bonn (LAB) survey of Galactic HI. Final data release of the combined LDS and IAR surveys with improved stray-radiation corrections. *Astron. Astrophys.* **440**, 775–782 (2005).



Extended Data Figure 1 | X-ray sources identified by Chandra in the central region of M82. **a**, Chandra image of the central region of M82 from the observation taken coincident with NuSTAR ObsID 006. The yellow circle shows the 70'' radius region used to extract NuSTAR fluxes. Within this region, 24 discrete X-ray point sources are identified, including X-1 and X-2.

b, Expanded view of the crowded central region of **a**. Yellow crosses indicate the locations of identified point sources. We have used, where possible, the numbering from ref. 13 (sources up to no. 15). After this we assign our own numerical identification (note that sources 4 and 10 from ref. 13 are not detected in this observation).



Extended Data Figure 2 | Swift imaging of the region containing M82 X-1 and M82 X-2. **a, b,** In greyscale are images obtained via Swift automated processing over the 5–10 keV band for all of the observations during early February (**a**; 2014 February 04 through 2014 February 11) and mid-March (**b**; 2014 March 7 through 2014 March 11). The early February observations have 68.5 ks of exposure (mostly because of the increased cadence of

observations due to the Swift monitoring of SN 2014J), while the mid-March snapshot has 1.8 ks of exposure. The images are 1.5 arcmin on a side and have been smoothed with a 2-pixel (4 arcsecond) Gaussian kernel (circle at lower left). The location of X-1 and X-2 are shown by the crosses in both panels. The late-time observation clearly shows a reduction in the flux from X-1 and that the flux is dominated by X-2.

Extended Data Table 1 | List of NuSTAR observations used in this analysis

ObsId	Date start (UT)	Date end (UT)	Effective Exposure (ks)
80002092002	2014-01-23 12:58:00	2014-01-24 23:01:05	65.818
80002092004	2014-01-25 19:53:58	2014-01-27 19:16:03	89.965
80002092006	2014-01-28 12:31:46	2014-02-04 05:29:48	309.525
80002092007	2014-02-04 06:07:53	2014-02-10 18:14:28	306.24
80002092008	2014-02-10 18:51:40	2014-02-11 12:00:24	33.648
80002092009	2014-02-11 12:37:59	2014-02-13 23:45:23	114.893
80002092011	2014-03-03 17:07:14	2014-03-06 01:02:13	110.925

Start and stop dates are given, as is the effective exposure of each observation. The exposure has been corrected for the period during which the source was occulted by the Earth, periods during which the instrument was not taking data and for the rate-dependent deadtime when the instrument was processing events and not sensitive to new incident photons.

Extended Data Table 2 | Best-fit orbital parameters

Parameter	Value
Orbital period P_{orb} (d)	2.53260(5)
Projected semi-major axis X (ls)	22.225(4)
Mid-eclipse time T_{90} (MJD)	56695.3659(1)

Orbital parameters determined by fitting a sinusoidal orbital modulation to ObsIDs 008–009. Error ranges for the parameters are given at the 1σ level.

Extended Data Table 3 | Best fit period and period derivatives for individual NuSTAR observations

ObsId	MJD	Period (s)	Period derivative (s/s)	Period 2nd derivative (s/s ²)	TOA Scatter (ms)
006	56685.5	1.3726001(4)	-3.0(1)x10 ⁻¹¹	0	62
007	56692.2	1.3725728(4)	8(3)x10 ⁻¹²	-4.2(1)x10 ⁻¹⁶	31
008-009	56698.7	1.3725076(4)	-1.38(7)x10 ⁻¹⁰	-3.3(5)x10 ⁻¹⁶	9
011	56719.8	1.3722225(6)	-2.73(7)x10 ⁻¹⁰	0	14

Pulse period and period derivatives fitted to time of arrival (TOA) data for individual observations after removal of the orbital modulation. Errors are 1σ, and the parentheses indicate the error on the last digit of the reported value.

Extended Data Table 4 | 3–10 keV flux contributions for bright sources near the nucleus of M82

Source	N_{H} (10^{22} cm^{-2})	Γ	3-10 keV flux contribution (%)
15	50^{+10}_{-20}	3 ± 1	4.4 ± 0.6
18	3 ± 1	0.4 ± 0.3	4.2 ± 0.6
9	6 ± 1	2.8 ± 0.4	1.3 ± 0.2
6	0.5 ± 0.3	1.6 ± 0.2	1.2 ± 0.2
20	2^{+4}_{-1}	< 1.0	$1.0^{+0.2}_{-0.4}$

Fluxes derived from power-law fits to the five brightest sources after M82 X-1 and X-2. Errors are 90% confidence. N_{H} is the hydrogen column and Γ is the power-law index.

Large, non-saturating magnetoresistance in WTe_2

Mazhar N. Ali¹, Jun Xiong², Steven Flynn¹, Jing Tao³, Quinn D. Gibson¹, Leslie M. Schoop¹, Tian Liang², Neel Haldolaarachchige¹, Max Hirschberger², N. P. Ong² & R. J. Cava¹

Magnetoresistance is the change in a material's electrical resistance in response to an applied magnetic field. Materials with large magnetoresistance have found use as magnetic sensors¹, in magnetic memory², and in hard drives³ at room temperature, and their rarity has motivated many fundamental studies in materials physics at low temperatures⁴. Here we report the observation of an extremely large positive magnetoresistance at low temperatures in the non-magnetic layered transition-metal dichalcogenide WTe_2 : 452,700 per cent at 4.5 kelvins in a magnetic field of 14.7 teslas, and 13 million per cent at 0.53 kelvins in a magnetic field of 60 teslas. In contrast with other materials, there is no saturation of the magnetoresistance value even at very high applied fields. Determination of the origin and consequences of this effect, and the fabrication of thin films, nanostructures and devices based on the extremely large positive magnetoresistance of WTe_2 , will represent a significant new direction in the study of magnetoresistivity.

Large magnetoresistance (MR) is an uncommon property, mostly of magnetic compounds. Giant magnetoresistance (GMR)⁵ and colossal magnetoresistance (CMR)^{6,7} occur in thin-film metals and manganese-based perovskites, for example. In contrast, ordinary magnetoresistance, a relatively weak effect, is commonly found in non-magnetic compounds and elements⁸. Magnetic materials typically have negative magnetoresistances (where MR, typically reported as a percentage, is defined as $[\rho(H) - \rho(0)]/\rho(0)$, and $\rho(H)$ is the resistivity in an applied magnetic field H). Positive magnetoresistance is seen in metals, semiconductors and semimetals⁹. It is usually at the level of a few per cent for metals; semiconducting silver chalcogenides have magnetoresistances of up to 12,000%, comparable with those of materials showing CMR¹⁰. For single-carrier-type semiconductors, the MR behaves as $(1 + \mu H^2)$, where μ is the carrier mobility; high-mobility semiconductors can therefore show relatively large effects¹¹. High-purity elemental bismuth, a semimetal, has an extremely large positive magnetoresistance, as does graphite¹². In semimetals, very high magnetoresistances are attributed to a balanced hole–electron ‘resonance’ condition; as described here, WTe_2 seems to be the first known material for which this resonance is nearly perfect.

WTe_2 is a layered transition-metal dichalcogenide (TMD) crystallizing in a distorted version of the common MoS_2 structure type¹³. TMDs are known to have many interesting properties, such as the catalysis of chemical reactions¹⁴, the presence of charge density waves (CDWs)¹⁵ and superconductivity¹⁶; they are solid-state lubricants¹⁷, make nanotubes¹⁸ and electrodes for rechargeable batteries¹⁹, have been exfoliated to fabricate state-of-the-art nanostructures^{20,21}, and now, we report, can show extremely large magnetoresistance. In the layered TMD compounds, metal layers are sandwiched between adjacent chalcogenide layers; this dichalcogenide sandwich stacks along the c axis of the hexagonal structure, with van der Waals bonding between layers. As a result of this anisotropic bonding, layered TMDs are typically electronically two-dimensional. In WTe_2 , however, there is an additional structural distortion: tungsten chains are formed within the dichalcogenide layers along the a axis of the orthorhombic unit cell, making the compound structurally one-dimensional (Fig. 1a). WTe_2 is a semimetal^{22,23} and has previously been

investigated for thermoelectric applications in solid solutions with WSe_2 and MoTe_2 (ref. 24).

We have discovered an extremely large positive magnetoresistance (XMR) in WTe_2 of up to 452,700% at 4.5 K in an applied field of 14.7 T when the current direction is along the tungsten chains (a axis) and the magnetic field is applied perpendicular to the dichalcogenide layers, along the c axis. The magnetoresistance is still increasing at 60 T, the highest field in our measurements, where it has a value of 13,000,000%. There is no indication of resistivity saturation even at these high applied fields, an indication of the unique character of WTe_2 . WTe_2 has a highly anisotropic electronic structure, with small pockets of holes and electrons (see below and ref. 23) in the directions in which tungsten chains are found in the crystal structure. The XMR is very anisotropic: with current flowing along the chain direction it is maximized when the field is applied perpendicular to the WTe_2 layers, and decreases by more than 90% when the magnetic field is applied in other directions. The effect becomes significant at temperatures below ~ 150 K; the temperature of the ‘turn on’ increases with the magnitude of the applied magnetic field. Electron diffraction patterns taken at low temperatures indicate

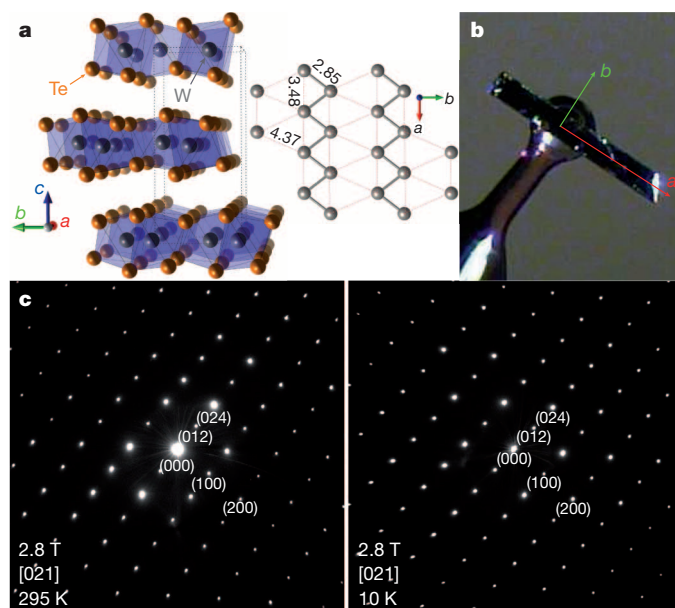


Figure 1 | Structural considerations. **a**, Crystal structure of WTe_2 , showing the layered nature, typical of TMDs, and also the chains of W atoms along the a axis distorting the ideal hexagonal net. All distances are in ångströms. **b**, A typical crystal of WTe_2 , with crystallographic directions marked. The XMR exists when the current (I) flows along a and the field is parallel to c (see below). **c**, Electron diffraction images looking down the $[021]$ zone showing the reciprocal lattice of WTe_2 at 295 and 10 K. The data, which show no superlattice formation on cooling, to a sensitivity of 1 part in 100,000, indicate that there is no structural transition in WTe_2 deep into the XMR regime. (The effective magnetic field at the sample in the TEM is about 2.8 T.)

¹Department of Chemistry, Princeton University, Princeton, New Jersey 08544, USA. ²Joseph Henry Laboratories and Department of Physics, Princeton University, Princeton, New Jersey 08544, USA.

³Department of Condensed Matter Physics and Materials Science, Brookhaven National Laboratory, Upton, New York 11973, USA.

that the origin of the observed effect is not linked to the onset of a charge density wave or a Peierls-like distortion, electronic instabilities that are frequently observed in layered TMDs.

The temperature-dependent resistivity under various applied magnetic fields ($\mu_0 H$ up to 14.7 T) is shown in Fig. 2. In zero field, the room-temperature resistivity is 0.6 m Ω cm and decreases to 1.9 $\mu\Omega$ cm by 2 K, yielding a residual resistivity ratio (see Methods) of ~ 370 . When a field is applied, the resistivity of the sample essentially follows the zero-field curve until it is cooled close to the 'turn on' temperature T^* (taken as the minimum in the resistivity) below which the resistivity begins to increase markedly. The magnetoresistance effect at low temperatures is extremely large, reaching 452,700% by 4.5 K in a field of 14.7 T. The 'turn on' temperature is shifted to a higher temperature (at the rate of ~ 4.4 K T^{-1} ; Fig. 2, upper inset) as larger fields are applied, implying competition between dominating scattering mechanisms. The transmission electron microscope (TEM) electron diffraction patterns taken at 10 K (Fig. 1c) in a ~ 2.8 T field (the field in the TEM at the sample position), well within the XMR phase regime, show no evidence of a superlattice to a sensitivity of about 1 part in 100,000 of the main diffraction peaks. This indicates that there is no structural phase transition or charge density wave accompanying the onset of the XMR effect.

Figure 3a shows the field dependence of the XMR at various temperatures. The upper inset shows the XMR at higher temperatures, and the lower inset shows the Shubnikov–de Haas quantum oscillations observed at low temperatures; the oscillations have been extracted after fitting a second-order polynomial to the 4.5 K parallel field measurement and subtracting that as background. They become visible by 6 T, begin increasing in amplitude, and then become dampened at about 10.5 T before re-emerging with much larger amplitude near 12.3 T. Figure 3b shows the dependence of the XMR effect on the angle of the applied field to the c axis. When the field is aligned parallel to the c axis the XMR effect is maximized, with $[\rho(14.7) - \rho(0)]/\rho(0) = 4,527$ (or 452,700%). As the field is rotated to align parallel to either the a axis or the b axis, the MR effect is greatly diminished and dies as the cosine of the angle; this large anisotropy is probably due to the very anisotropic Fermi surface of WTe_2 and scattering rates. Measurements up to 60 T at 0.53 K show an XMR of $[\rho(60) - \rho(0)]/\rho(0) \approx 130,000$ (MR $\approx 13,000,000\%$) with strong quantum oscillations and still no resistivity saturation (Fig. 3c). The dependence of the resistivity on magnetic field is close to quadratic

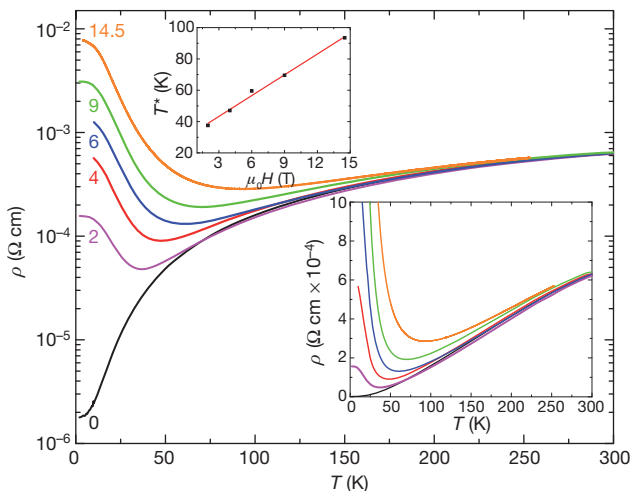


Figure 2 | The temperature and field dependence of the XMR in WTe_2 , for I parallel to a and H parallel to c . Plots of resistivity against temperature; numbers against graphs are field strengths in teslas. Lower inset: plot of resistivity against temperature, showing the effect being turned on. T^* is defined as the temperature at which the resistivity is a minimum—an approximation of the temperature at which the XMR is turned on. Upper inset: linear dependence of T^* on magnetic field; the slope is 4.4 K T^{-1} .

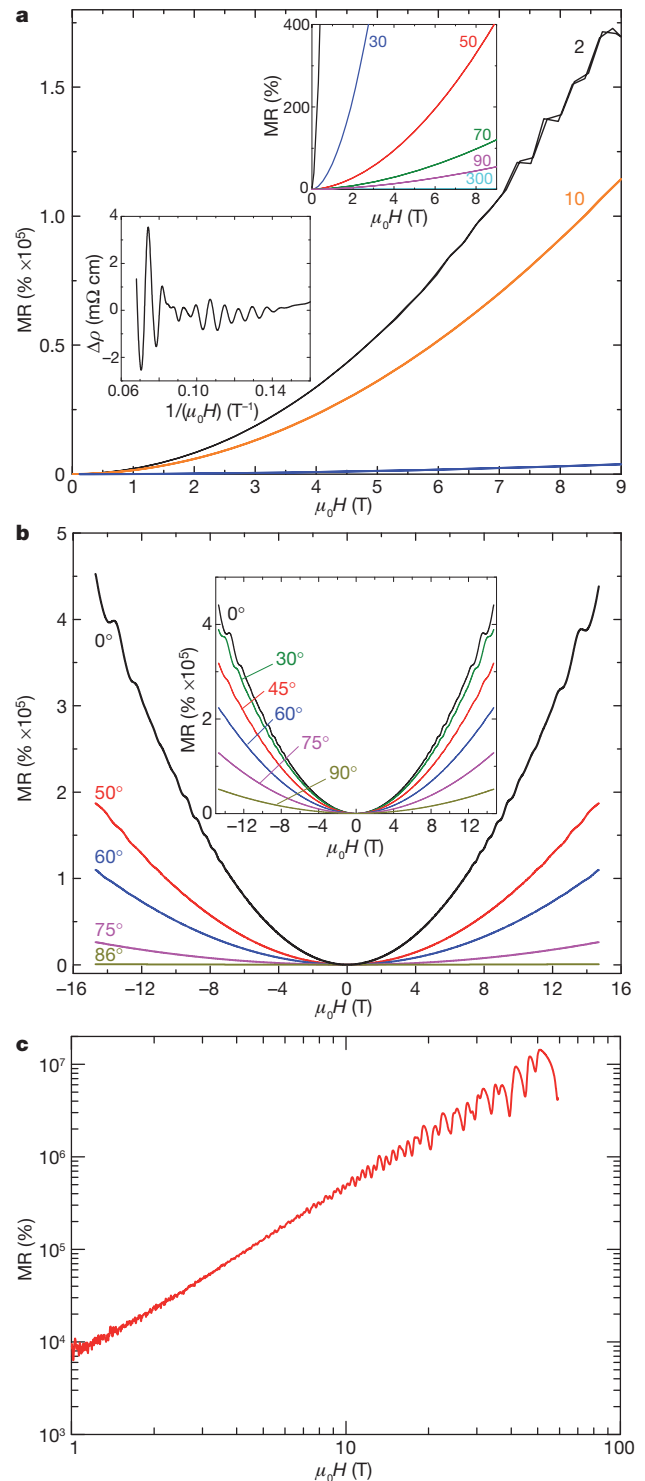


Figure 3 | Field and angular dependence of the XMR in WTe_2 . **a**, Field dependence of the XMR in WTe_2 with the current along the W – W chains (a axis) and the applied field parallel to the c axis from 0 to 9 T at representative temperatures (numbers in kelvins). Upper inset: detail of the magnetoresistance at higher temperatures. Lower inset: detail of the quantum oscillations at 4.5 K. This demonstrates the high quality of the crystal. **b**, Angular dependence of the XMR at 4.5 K with the current parallel to a , showing that the effect is one-dimensional; it is maximized at 0° (when H is parallel to c) and goes to 0 at 90° (when H is perpendicular to c). The main panel shows the MR as the applied field is rotated to be parallel to a , and the inset shows the same effect when the field is rotated to be parallel to b . **c**, XMR of WTe_2 up to 60 T at 0.53 K, with I parallel to a and H parallel to c , for $\rho(H)/\rho(0) \propto H^m$ ($m \approx 2$).

even up to the highest fields measured, with no indication of a change in dependence.

Our electronic structure calculations, similar to those in a previous report²³, show WTe_2 to be a semimetal. The valence band and conduction bands barely cross the Fermi energy at different places in the Brillouin zone (Fig. 4a), in an electronic structure that is reminiscent of that of the excitonic insulator TiSe_2 (ref. 25). The small electron and hole pockets along the Γ - X direction (corresponding to the a axis, along the tungsten chains in real space) are what make WTe_2 a semimetal. The detailed shape of the Fermi surface (Fig. 4b) is very sensitive to the position of the Fermi level. A potential second set of electron and hole pockets forming along Z - U (parallel to the Γ - X direction, but shifted along k_z into the perpendicular face of the Brillouin zone) represents a potential second crossing that would change the pockets into tubes in the Fermi surface. Future study by angle-resolved photoemission spectroscopy, transport analysis and characterization of the quantum oscillations will be needed to determine the details of the Fermi surface shape and character, and its link to the observed XMR.

With a positive magnetoresistance as large as this, there are few systems comparable to WTe_2 . High-purity graphite and bismuth display XMR at low T . It has been shown²⁶ that in graphite this behaviour is well described by a two-band model for charge transport in semimetals, where n -type and p -type carriers are both present; the analysis, however, is restricted to very low H (200 mT or less). As H increases, the MR in graphite and Bi deviate markedly from the modelled H^2 dependence, and eventually saturate at large H (refs 27, 28). Both the saturating MR in graphite and Bi and the non-saturating MR in WTe_2 are consequences of a resonance in semimetal charge transport, which indicates that the MR peaks at $p/n = 1$ and that the stringency of the $p/n = 1$ resonance requirement becomes more and more severe with increasing H (see Methods). Hence, as long as resonance is maintained, the resistivity increases as H^2 without saturation; however, a slight deviation from perfect n - p compensation will cause the resistivity to saturate to a field-independent value. This occurs in both graphite and Bi, but WTe_2 seems to be the first material known that displays a non-saturating MR and thus nearly perfect n - p compensation. WTe_2 has an extremely small overlap between valence-band and conduction-band states²³, and

in that way it resembles an excitonic insulator²⁵. It may be that this excitonic-insulator-like character of WTe_2 , in which each electron has a corresponding hole, leads to the nearly perfectly balanced electron-hole populations that are responsible for the XMR effect that does not saturate at very high fields; further work will be required to verify this.

Particularly important differences between Bi and WTe_2 are seen from the materials perspective. For the development of advanced devices, in addition to the fact that ordinary purity WTe_2 shows the XMR effect (we observed it in WTe_2 synthesized from 99.9% W and 99.99% Te; in contrast, 99.9995% Bi shows the very large MR²⁶), the fact that WTe_2 is a layered TMD that can be easily exfoliated allows it to be used as the basis for thin films and advanced nanostructure devices similar to those being actively pursued that are based on MoS_2 (ref. 21) and related TMDs. It is especially important that WTe_2 is a binary compound. Elemental Bi has few avenues for tuning chemically, whereas WTe_2 belongs to a materials family that is well known to be tunable by chemical doping. In addition to tuning by W and Te site doping, intercalation reactions accompanied by charge transfer to the MX_2 host, expected to be possible for many elements in the periodic table including magnetic ones, are common in layered TMDs, and the same will be possible for WTe_2 .

The single crystals made in this study were crudely exfoliated with double-sided tape, and thicknesses down to a few micrometres were easily achieved. Evaporation, chemical vapour deposition or laser ablation growth and subsequent annealing to make single-crystal thin films may further enhance the MR effect. Hybrid structures of various kinds, including the layering of WTe_2 with magnetic films, may be useful in devices such as highly sensitive low-temperature magnetic-field sensors or high-field temperature sensors in cryogenics. In particular, the one-dimensional aspect of the XMR in WTe_2 may be useful in low-temperature magnetic-field sensing and, especially, orienting. In fact, it was reported recently that below 20 K or above 5 T (refs 29, 30) the materials currently used for temperature or field measurements are prone to large degrees of error. In contrast, this regime is where WTe_2 performs best. The ease with which this system can be exfoliated, in addition to the fact that even small changes in the electron or hole concentration should change the Fermi surface—and thus, it can be expected, the XMR—make WTe_2 an ideal candidate for electron-gating experiments. It may

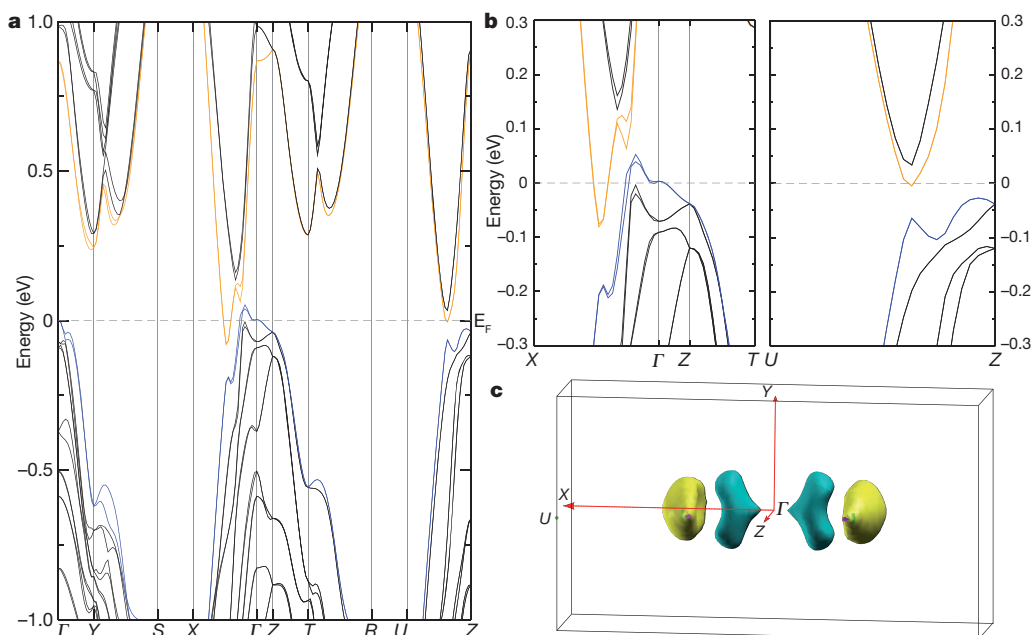


Figure 4 | The electronic structure of WTe_2 , calculated including spin orbit coupling. **a**, The energy-wavevector relationships for high-symmetry directions in the orthorhombic Brillouin zone. Note the semimetal character due to the crossing of the hole and electron bands near Γ . E_F , Fermi energy.

b, Detail of the calculated electronic structure in the Γ - X and Γ - Z directions and similar detail showing the possible crossing of the states between Z and U . **c**, The Fermi surface of WTe_2 showing electron (yellow) and hole (blue) pockets displaced from the Γ point and aligned along the chain direction.

be possible to make nanotubes from WTe_2 , opening even more possibilities for nanostructure device fabrication. Careful chemical or electronic doping of WTe_2 may be crucial in elucidating the cause of the XMR and potentially unlocking further properties of interest.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 May; accepted 28 July 2014.

Published online 14 September 2014.

1. Lenz, J. E. A review of magnetic sensors. *Proc. IEEE* **78**, 973–989 (1990).
2. Moritomo, Y., Asamitsu, A., Kuwahara, H. & Tokura, Y. Giant magnetoresistance of manganese oxides with a layered perovskite structure. *Nature* **380**, 141–144 (1996).
3. Daughton, J. GMR applications. *J. Magn. Magn. Mater.* **192**, 334–342 (1999).
4. Urushibara, A. *et al.* Insulator–metal transition and giant magnetoresistance in $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$. *Phys. Rev. B* **51**, 14103–14109 (1995).
5. Egelhoff, W. F. *et al.* Magnetoresistance values exceeding 21% in symmetric spin valves. *J. Appl. Phys.* **78**, 273–277 (1995).
6. Ramirez, A. P., Cava, R. J. & Krajewski, J. Colossal magnetoresistance in Cr-based chalcogenide spinels. *Nature* **386**, 156–159 (1997).
7. Jin, S., McCormack, M., Tiefel, T. H. & Ramesh, R. Colossal magnetoresistance in LaCaMnO ferromagnetic thin films. *J. Appl. Phys.* **76**, 6929–6933 (1994).
8. Yang, F. Y. *et al.* Large magnetoresistance of electrodeposited single-crystal bismuth thin films. *Science* **284**, 1335–1337 (1999).
9. Mun, E. *et al.* Magnetic field effects on transport properties of PtSn_4 . *Phys. Rev. B* **85**, 035135 (2012).
10. Ishiwata, S. *et al.* Extremely high electron mobility in a phonon-glass semimetal. *Nature Mater.* **12**, 512–517 (2013).
11. Solin, S. A., Thio, T., Hines, D. R. & Heremans, J. J. Enhanced room-temperature geometric magnetoresistance in inhomogeneous narrow-gap semiconductors. *Science* **289**, 1530–1532 (2000).
12. Alers, P. B. & Webber, R. T. The magnetoresistance of bismuth crystals at low temperatures. *Phys. Rev.* **91**, 1060–1065 (1953).
13. Brown, B. E. The crystal structures of WTe_2 and high-temperature MoTe_2 . *Acta Crystallogr.* **20**, 268–274 (1966).
14. Li, Y. *et al.* MoS_2 nanoparticles grown on graphene: an advanced catalyst for the hydrogen evolution reaction. *J. Am. Chem. Soc.* **133**, 7296–7299 (2011).
15. Moncton, D. E., Axe, J. D. & DiSalvo, F. J. Neutron scattering study of the charge-density wave transitions in 2H-TaSe_2 and 2H-NbSe_2 . *Phys. Rev. B* **16**, 801–819 (1977).
16. Morris, R. C., Coleman, R. V. & Bhandari, R. Superconductivity and magnetoresistance in NbSe_2 . *Phys. Rev. B* **5**, 895–901 (1972).
17. Rapoport, L. *et al.* Hollow nanoparticles of WS_2 as potential solid-state lubricants. *Nature* **387**, 791–793 (1997).
18. Rapoport, L., Fleischer, N. & Tenne, R. Applications of WS_2 (MoS_2) inorganic nanotubes and fullerene-like nanoparticles for solid lubrication and for structural nanocomposites. *J. Mater. Chem.* **15**, 1782–1788 (2005).
19. Bates, J., Gruzalski, G., Dudney, N., Luck, C. & Yu, X. Rechargeable thin-film lithium batteries. *Solid State Ion.* **70**, 619–628 (1994).
20. Ayari, A., Cobas, E., Ogundadegbe, O. & Fuhrer, M. S. Realization and electrical characterization of ultrathin crystals of layered transition-metal dichalcogenides. *J. Appl. Phys.* **101**, 014507 (2007).
21. Radisavljevic, B., Radenovic, A., Brivio, J., Giacometti, V. & Kis, A. Single-layer MoS_2 transistors. *Nature Nanotechnol.* **6**, 147–150 (2011).
22. Kabashima, S. Electrical properties of tungsten-ditelluride WTe_2 . *J. Phys. Soc. Jpn.* **21**, 945–948 (1966).
23. Augustin, J. *et al.* Electronic band structure of the layered compound Td-WTe_2 . *Phys. Rev. B* **62**, 10812–10823 (2000).
24. Revolinsky, E. & Beerntsen, D. Electrical properties of the MoTe_2 – WTe_2 and MoSe_2 – WSe_2 systems. *J. Appl. Phys.* **35**, 2086–2089 (1964).
25. Pillo, T. *et al.* Photoemission of bands above the Fermi level: the excitonic insulator phase transition in 1T-TiSe_2 . *Phys. Rev. B* **61**, 16213–16222 (2000).
26. Du, X., Tsai, S.-W., Maslov, D. L. & Hebard, A. F. Metal-insulator-like behavior in semimetallic bismuth and graphite. *Phys. Rev. Lett.* **94**, 166601 (2005).
27. Fauqué, B., Vignolle, B., Proust, C., Issi, J.-P. & Behnia, K. Electronic instability in bismuth far beyond the quantum limit. *New J. Phys.* **11**, 113012 (2009).
28. Kopelevich, Y. *et al.* Reentrant metallic behavior of graphite in the quantum limit. *Phys. Rev. Lett.* **90**, 156402 (2003).
29. Ekin, J. *Experimental Techniques for Low-Temperature Measurements: Cryostat Design, Material Properties and Superconductor Critical-Current Testing* (Oxford Univ. Press, 2006).
30. Jakub Jankowski, S. E.-A. & Oszwaldowski, M. Hall sensors for extreme temperatures. *Sensors* **11**, 876–885 (2011).

Acknowledgements We thank T. Valla, I. Pletikoscic, F. Balakirev, R. McDonald and J. Betts for discussions, and E. Tutuc for inquiring about WTe_2 . This research was supported by the Army Research Office, grants W911NF-12-1-0461 and W911NF-11-1-0379, and the NSF MRSEC Program Grant DMR-0819860. The National Magnet Laboratory is supported by the National Science Foundation Cooperative Agreement no. DMR-1157490, the State of Florida, and the US Department of Energy; this work was supported by the US Department of Energy's Basic Energy Sciences (DOE BES) project 'Science at 100 Tesla'. The electron microscopy study at Brookhaven National Laboratory was supported by the DOE BES, by the Materials Sciences and Engineering Division under contract DE-AC02-98CH10886, and through the use of the Center for Functional Nanomaterials.

Author Contributions M.N.A. was the lead researcher. M.N.A. and S.F. grew crystals and measured resistivities with N.H. J.X., T.L. and M.H. measured the detailed resistivity behaviour. Q.D.G. and L.S. calculated the electronic structure. J.T. performed the electron microscopy characterization. N.P.O. and R.J.C. supervised the research. All authors contributed to writing the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.N.A. (mnali@princeton.edu).

METHODS

High-quality single crystals of WTe_2 were grown by means of Br_2 vapour transport. Tungsten powder (99.9%) was ground together with purified tellurium (99.99%), pressed into a pellet and heated in an evacuated quartz tube at 700°C for 2 days; this was homogenized and then reheated at 750°C for a further 2 days. This final pellet was ground into a fine powder and a temperature gradient of 100°C between 750 and 650°C in a three-zone furnace was used for crystal growth, with a Br_2 concentration of $\sim 3\text{ mg ml}^{-1}$ in a sealed quartz tube for 1 week. Optimal crystals were obtained under these conditions; crystals grown at higher temperatures showed substantially lower residual resistivity ratios $\rho(300\text{ K})/\rho(2\text{ K})$ and degraded magnetoresistance. The need to employ low-temperature synthesis to avoid defect formation that degrades properties is frequently observed in TMDs, for example in TiSe_2 (ref. 31). The crystals grew as thin ribbons (Fig. 1b), with the long direction being the $\text{W}-\text{W}$ chain direction and the larger flat faces being perpendicular to the stacking direction of the layers.

WTe_2 crystals were structurally and chemically characterized by powder X-ray diffraction to confirm bulk purity, single-crystal X-ray diffraction to determine crystal growth orientation, scanning electron microscopy with energy dispersive X-ray spectroscopy for chemical analysis, and TEM to search for a low-temperature phase transition. For general electronic characterization, superconducting quantum interference device magnetometer measurements revealed weak diamagnetism, and thermopower measurements confirmed a previously reported n - p -type crossover at 65 K in a 0 T field and n -type Hall effect behaviour down to 2 K (ref. 22). In resistivity measurements, electrical anisotropy was found; the tungsten chain direction (the a axis) had the lowest resistivity.

Powder X-ray diffraction patterns were collected using $\text{Cu K}\alpha$ radiation on a Bruker D8 Focus diffractometer with a graphite monochromator. Single-crystal X-ray diffraction data were collected on a Bruker APEX II using $\text{Mo K}\alpha$ radiation ($\lambda = 0.71073\text{ \AA}$) at 100 K . Scanning electron microscopy and energy dispersive X-ray analysis were conducted on an FEI Quanta 200 FEG Environmental SEM and were used to determine the composition of the crystals. Electron diffraction was performed in a Jeol 3000F transmission electron microscope equipped with a Gatan liquid-helium cooling stage.

The magnetoresistance of WTe_2 samples was measured with the four-point probe method in a Quantum Design PPMS and with a Delta-mode method by a Keithley 6221 current source meter and a 2182A nanovoltmeter. The high-field dependent data were taken at 4.5 K up to 14.7 T in an American Magnetics superconducting magnet. Resistivity measurements up to 60 T in a pulsed field were performed at the National High Magnetic Field Laboratory.

The electronic structure calculations were performed in the framework of density functional theory, using the WIEN2K³² code with a full-potential linearized augmented plane-wave and local orbitals (FP-LAPW + lo) basis³³ together with the Perdew–Burke–Ernzerhof parameterization of the generalized gradient approximation as the exchange–correlation functional. The plane wave cut-off parameter $R_{\text{MT}}K_{\text{max}}$ was set to 8 and the Brillouin zone was sampled by 10,000 k points. The convergence was confirmed by increasing both $R_{\text{MT}}K_{\text{max}}$ and the number of k points until no change in the total energy was observed. The Fermi surface was plotted with the program Xcrysden.

More detail on the magnetoresistance of semimetals and comparison of the behaviour of WTe_2 and Bi is provided below.

MR is defined as $(\rho_{xx}(H) - \rho_0)/\rho_0$, with $\rho_{xx}(H)$ the resistivity in a magnetic field H and $\rho_0 = \rho_{xx}(0)$. Semimetals, for example graphite and bismuth, display very large MR at low temperatures T because compensation leads to near-cancellation of the Hall E field. In weak H , the MR increases rapidly, as $(\mu H)^2$, with μ the carrier mobility. Carrier compensation would therefore seem to be a promising way to attain colossal MR, but so far the MR in semimetals deviates from the H^2 form at

large H to saturate to a constant value. The remarkable feature of the MR in WTe_2 is that it maintains its H^2 behaviour to reach 13,000,000% at 60 T . We discuss here how a sharp ‘resonance’ in semimetals can enable extremely large MR to persist to very large H without saturation.

Semiclassically (ignoring quantum oscillations associated with Landau levels), ρ_{xx} displays a remarkably narrow resonance as p/n is varied with B fixed, regardless of the mobility ratio μ'/μ (Extended Data Fig. 1). Here, n and μ are the density and mobility of the electrons, and p and μ' are the corresponding quantities for holes ($B = \mu_0 H$, where μ_0 is the free-space permeability). At the peak ($p/n = 1$), the MR equals $(\bar{\mu}B)^2$, with $\bar{\mu} = \sqrt{\mu\mu'}$. Hence, as long as resonance is maintained, the MR increases as B^2 without saturation, at a rate dictated by $\bar{\mu}$. However, because the Q factor of the resonance grows linearly with B , this is an increasingly stringent condition when $\bar{\mu}B \gg 1$. A slight deviation from perfect compensation will force ρ_{xx} off resonance and cause it to saturate to a field-independent value. This seems to be true for both graphite and Bi. In Bi, an additional complication is the extreme velocity anisotropy of the Dirac pockets, which require treatments beyond the relaxation-time approximation. The challenge then is to remain on resonance to very large H . The MR in WTe_2 continues to increase as H^a , with $a \approx 2$ up to 60 T (or $\bar{\mu}B = 325$, using $\bar{\mu} \approx 5$). This seems to be the first example uncovered of non-saturating MR.

In the semiclassical two-band isotropic model, the total conductivity tensor is conveniently expressed in the complex representation

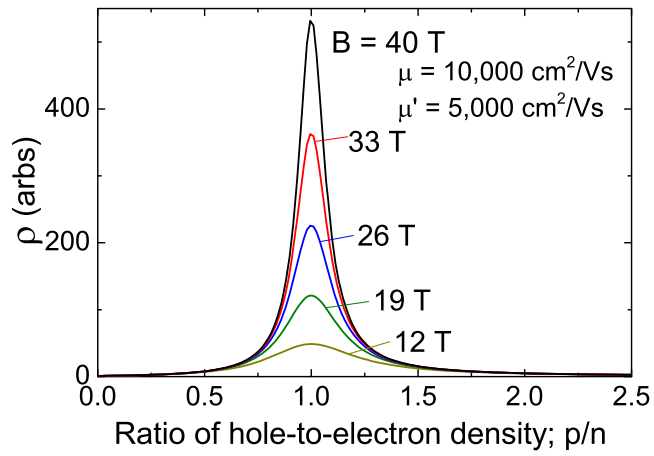
$$\hat{\sigma} = e \left[\frac{n\mu}{(1 + i\mu B)} + \frac{p\mu'}{(1 - i\mu' B)} \right]$$

where $e > 0$ is the charge, and μ and μ' are the mobilities of electrons and holes, respectively. The complex resistivity is the reciprocal of $\hat{\sigma}$, namely

$$\hat{\rho} = \frac{1 + \mu\mu'B^2 + i(\mu - \mu')B}{e(n\mu + p\mu' + i(p - n)\mu\mu'B)}. \quad (1)$$

From equation (1), the experimentally observed resistivity ρ_{xx} equals $\text{Re } \hat{\rho}$ (the Hall resistivity $\rho_{xy} = \text{Im } \hat{\rho}$). ρ_{xx} has a Lorentzian form that peaks sharply at $p = n$, irrespective of the ratio μ/μ' . Extended Data Figure 1 shows the resonant behaviour of ρ_{xx} against p/n with B fixed at selected values (μ/μ' is fixed at 2). The width of the resonance is obtained by setting $(p - n)\mu\mu'B/(n\mu + p\mu')$ to 1. We then have for the Q factor of the resonance $Q = \langle\mu\rangle B$, with the average mobility $\langle\mu\rangle = (1/\mu + 1/\mu')^{-1}$. The curves in Extended Data Fig. 1 verify that Q increases in proportion to B . The resonant form implies that at exact compensation ($n = p$), ρ_{xx} grows as B^2 without limit with an MR given by $\Delta\rho/\rho = \bar{\mu}^2 B^2$. However, if p/n deviates from 1, the system falls off resonance and ρ_{xx} saturates to the field-independent value $\rho_\infty = (n\mu' + p\mu)/e[(p - n)^2\mu\mu']$ in the limit $\bar{\mu}B \gg 1$. As Q increases with B , the saturation occurs at very small deviation of p/n from 1 at large B . Conversely, in weak B , an MR with B^2 variation is easily achieved even when n and p are only approximately equal, because the resonance is broad (see curve at $B = 12$). This is the situation in high-purity graphite for $B < 0.2\text{ T}$ analysed previously²⁶. When B exceeds a few teslas, strong deviation of ρ_{xx} becomes apparent²⁸. Similarly in high-purity Bi, ρ_{xx} deviates from B^2 above 1 T and approaches saturation²⁷ above 10 T .

- Di Salvo, F. J., Moncton, D. E. & Waszczak, J. V. Electronic properties and superlattice formation in the semimetal TiSe_2 . *Phys. Rev. B* **14**, 4321–4328 (1976).
- Blaha, P., Schwarz, K., Madsen, G., Kvasnicka, D. & Luitz, J. *WIEN2k, an Augmented Plane Wave + Local Orbitals Program for calculating Crystal Properties* (Technische Univ. Wien, 2001).
- Singh, D. J. & Nordström, L. *Planewaves, Pseudopotentials, and the LAPW Method* 2nd edn (Springer, 2006).



Extended Data Figure 1 | Simulation of the resonance criterion for non-saturating MR in semimetals. The MR is sharply peaked at a ratio of holes to electrons of 1:1, even when their mobilities are not equal, and the MR continues to increase to high applied fields with no saturation.

Vibrational spectroscopy in the electron microscope

Ondrej L. Krivanek^{1,2}, Tracy C. Lovejoy¹, Niklas Dellby¹, Toshihiro Aoki³, R. W. Carpenter⁴, Peter Rez², Emmanuel Soignard^{3,4}, Jiangtao Zhu^{3,†}, Philip E. Batson^{5,6}, Maureen J. Lagos^{5,6}, Ray F. Egerton⁷ & Peter A. Crozier⁸

Vibrational spectroscopies using infrared radiation^{1,2}, Raman scattering³, neutrons⁴, low-energy electrons⁵ and inelastic electron tunneling⁶ are powerful techniques that can analyse bonding arrangements, identify chemical compounds and probe many other important properties of materials. The spatial resolution of these spectroscopies is typically one micrometre or more, although it can reach a few tens of nanometres or even a few ångströms when enhanced by the presence of a sharp metallic tip^{6,7}. If vibrational spectroscopy could be combined with the spatial resolution and flexibility of the transmission electron microscope, it would open up the study of vibrational modes in many different types of nanostructures. Unfortunately, the energy resolution of electron energy loss spectroscopy performed in the electron microscope has until now been too poor to allow such a combination. Recent developments that have improved the attainable energy resolution of electron energy loss spectroscopy in a scanning transmission electron microscope to around ten millielectronvolts now allow vibrational spectroscopy to be carried out in the electron microscope. Here we describe the innovations responsible for the progress, and present examples of applications in inorganic and organic materials, including the detection of hydrogen. We also demonstrate that the vibrational signal has both high- and low-spatial-resolution components, that the first component can be used to map vibrational features at nanometre-level resolution, and that the second component can be used for analysis carried out with the beam positioned just outside the sample—that is, for ‘aloof’ spectroscopy that largely avoids radiation damage.

In the past two decades, the performance of electron microscopes has been greatly improved by the introduction of multipole-based aberration correction technology^{8–11}. Spatial resolution of better than 0.5 Å is now possible¹² and 1 Å is routine. In the scanning transmission electron microscope (STEM), aberration-corrected optics have also allowed the electron current in an atom-sized probe to be increased by 10× and more. This has led to new performance standards, such as non-destructive resolution and identification of all the atoms in a monolayer sample containing various impurities¹³, and electron energy loss spectroscopy (EELS) fine structure studies of individual atoms^{14–16}. These successes have given rise to a question: can similar technologies improve the energy resolution of EELS carried out in the electron microscope so that vibrational spectra with key features occurring at 50–500 meV become accessible at high spatial resolution, in a wide variety of samples and sample geometries?

Because of recent progress, we are now able to answer the question in the positive. The progress has taken place on three principal fronts: (1) the energy resolution of EELS carried out in the electron microscope has been improved to around 10 meV; (2) the EELS–STEM instrument has been optimized so that the electron probe incident on the sample contains a current sufficient to perform EELS experiments even when the energy width of the probe is ~10 meV and its size <1 nm; and (3) the tail of the intense zero loss peak (ZLP) in the EELS spectrum has been reduced so that it does not obscure the vibrational features of interest.

The innovations responsible for the progress are (1) a monochromator of a new design¹⁷, which is able to reach an energy resolution comparable to the highest resolution attained previously^{18,19} in broad-beam systems, and which allows an atom-sized electron probe to be formed; (2) an ultra-bright cold field emission electron gun, a dispersing–undispersing monochromator design, and aberration-corrected optics, which together maximize the intensity of the monochromated electron probe; and (3) increasing the energy dispersion of the spectrometer to 1 meV per channel (and higher), and developing operating modes that minimize ZLP tails due to aberrations and stray scattering. A fuller description of the innovations is provided in Methods. Recent improvements in the stability of the whole system are illustrated in Extended Data Fig. 1.

Figure 1 illustrates how the above developments are revealing spectral features that have long been ‘hidden in plain sight’ in electron microscopes—obscured by a wide ZLP. Figure 1a shows a ZLP obtained with the monochromator slit nearly closed, with a full-width at half-maximum (FWHM) of 9 meV, compared to an unmonochromated spectrum obtained with the system’s energy-selecting slit withdrawn from the beam. Figure 1b shows a spectrum obtained from hexagonal boron nitride (h-BN) with the slit slightly more open, to admit a larger current while keeping the probe size small, at two vertical scales: normalized so that the intensity maximum of the ZLP is equal to 1, and multiplied by 1,000, to show the region of the spectrum from 100 to 200 meV. The spectrum was obtained with an electron beam of about 10 pA current and 1 nm diameter. The strong feature at 173 meV corresponds to the energy of the longitudinal optical (LO) phonon in h-BN at 175 meV (see Methods).

Figure 2 shows vibrational excitations recorded in a range of materials: h-BN, SiO₂, SiC, TiH₂ and epoxy resin. The spectra were acquired using various electron-optical settings and total acquisition times ranging from 5 s to 50 s, as described in Methods. For the TiH₂ spectrum, the beam was about 5 nm outside the sample, in a so-called aloof position, to minimize radiation damage. The resin spectrum was recorded in an angle-resolved mode (see Methods) and came from a sample area of about 400 nm in diameter. The intensity of the vibrational peaks is of the order of 3×10^{-5} to 5×10^{-4} (above background) relative to the ZLP. All the peak energies observed here correspond closely to energies recorded by other vibrational spectroscopies: infrared, Raman, low-energy electron and neutron (as shown in Methods and Extended Data Fig. 2).

The observation of vibrational peaks due to hydrogen in TiH₂ and in the epoxy resin is especially interesting. In TiH₂, hydrogen is mobile and bound only weakly, which results in the relatively low (for hydrogen) vibrational energy of 147 meV. In epoxy resin, hydrogen is mostly bound to carbon, and 360 meV ($2,900 \text{ cm}^{-1}$) is a typical C–H stretch vibrational energy^{1–3}. Up to now, hydrogen has been essentially invisible in electron microscopes, its presence typically inferred from the modified electron distribution due to the electron it contributes to the sample’s electron density distribution. Its unambiguous detection by vibrational spectroscopy promises to provide a general technique for hydrogen

¹Nion Company, 1102 Eighth Street, Kirkland, Washington 98033, USA. ²Department of Physics, Arizona State University, Tempe, Arizona 85287, USA. ³LeRoy Eyring Center for Solid State Science, Arizona State University, Tempe, Arizona 85287, USA. ⁴Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona 85287, USA. ⁵Institute for Advanced Materials, Devices and Nanotechnology, Rutgers University, Piscataway, New Jersey 08854, USA. ⁶Departments of Physics and Materials Science, Rutgers University, Piscataway, New Jersey 08854, USA. ⁷Department of Physics, University of Alberta, Edmonton T6G 2E1, Canada. ⁸School of Engineering of Matter, Transport and Energy, Arizona State University, Tempe, Arizona 85287, USA. [†]Present address: TDK Headway Technologies Incorporated, Milpitas, California 95035, USA.

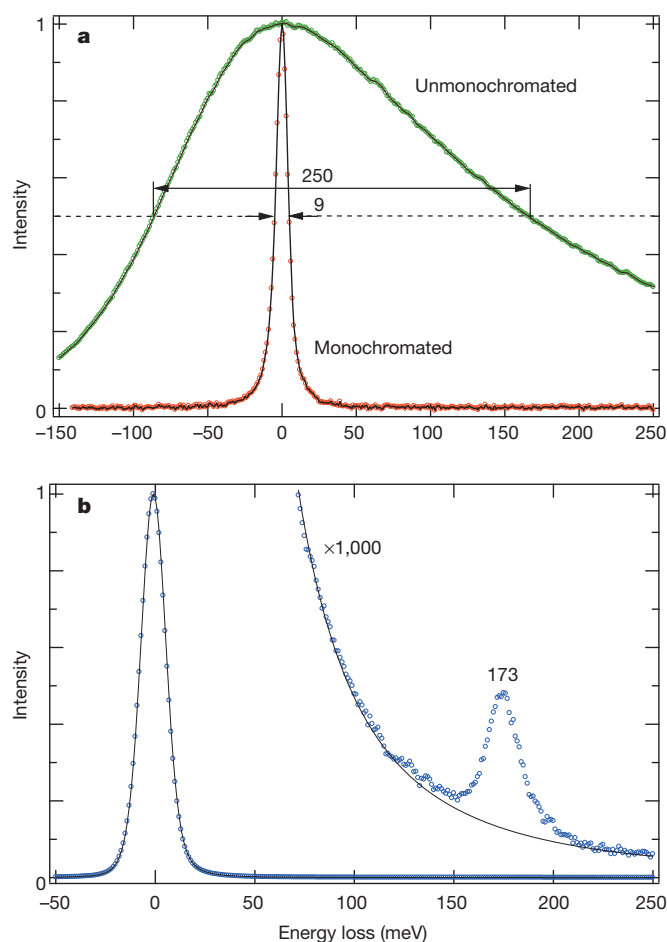


Figure 1 | Revealing vibrational signals in the electron microscope. **a**, 9 meV wide (full-width at half-maximum, FWHM) monochromated zero loss peak (ZLP) compared to the energy distribution of an unmonochromated beam produced by the system's cold field emission electron gun. **b**, 14 meV wide ZLP obtained with the energy-selecting slit more open (left), and 18 meV wide LO phonon in h-BN, with a peak energy of 173 meV (right).

detection in the many hydrogen-containing materials studied by electron microscopy. Another attractive prospect involves analysing the types of covalent hydrogen bonding present in microscopic amounts of matter, with H-C, H-N, H-O and other types of hydrogen bonds giving distinct vibrational frequencies^{1–3}.

The width of the vibrational peaks shown in Fig. 2 is 18–40 meV, which is worse than the resolution attainable by other vibrational spectroscopies^{1–7}. (The causes of the broadening are discussed in Methods.) However, the main interest of vibrational spectroscopy in an electron microscope lies in studying the spatial variation of vibrational modes in different nanostructures, ideally with near-atomic resolution. With the present energy resolution, the vibrational signal can be distinguished clearly enough to study the spatial variation, and the energy resolution is almost certain to improve further in the future.

Figure 3 explores the spatial resolution of our technique, by imaging the variation of the SiO₂ optical phonon at Si–SiO₂ and SiO₂–vacuum interfaces. A probe of about 2-nm diameter was advanced in 2-nm steps, and a single 10-s spectrum was recorded at each new position. A second scan was performed over the same locations to record the variation of the unsaturated ZLP, and a third scan to record medium-energy EELS losses up to 250 eV, whose strength was used to determine the sample thickness, also shown in the figure. To make sure that the sample did not shift significantly during the profile acquisition and that no significant radiation damage occurred, the examined area was imaged both before and after the spectroscopic line scans, using the high-angle annular dark field (HAADF) signal.

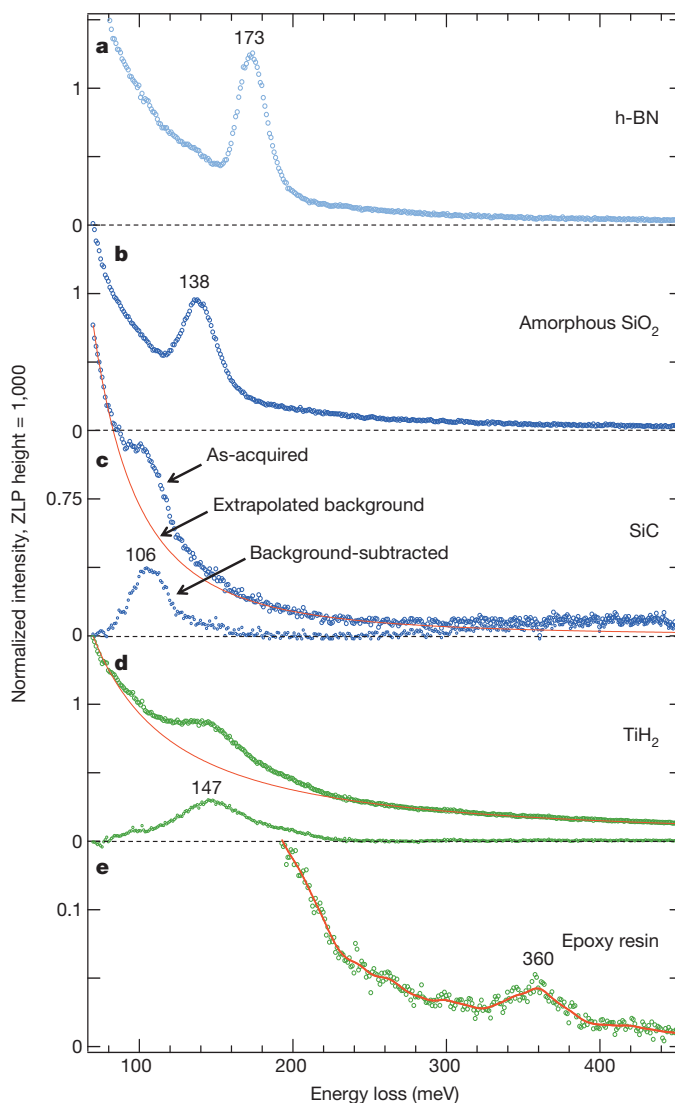


Figure 2 | Vibrational spectra from various materials. **a**, h-BN; **b**, amorphous SiO₂; **c**, SiC; **d**, TiH₂; and **e**, epoxy resin.

The vibrational signal was obtained by subtracting the background under the peak at 138 meV in all the spectra, which were similar to the spectrum shown in Fig. 2b. The signal was then normalized by dividing it by the intensity of the ZLP. It is seen to extend into the vacuum beyond the edge of the SiO₂ layer, and to decay to 1/e intensity about 40 nm from the edge. The HAADF image and the medium-energy EELS line profile show no such extended feature. This confirms that the spatially extended signal is due to optical phonons being excited even when the electron beam is passing tens of nanometres away from the sample.

Aloof beams losing energy to delocalized electronic excitations (such as surface plasmons) have been studied extensively in low-loss EELS^{20–22}. The mechanism can be described by dielectric response theory^{23–25}. The probability of excitation varies as $K_0[(2\omega R)/(\gamma v)]$ with the distance R from the specimen (see Methods and Extended Data Fig. 3), where K_0 is a modified Bessel function of the second kind, ω is the excitation frequency, v is the electron velocity and γ is the relativistic (Lorentz) factor. For optical phonon energies of 100–200 meV, this function shows tails extending tens and even hundreds of nanometres with an exponential decay, in good agreement with the present observations.

The spatial resolution obtainable with the aloof signal is comparable to that of tip-enhanced optical spectroscopy⁷, without needing to have a sharp tip in the vicinity of the examined structure. Because the interaction distance for the signal can be much larger than the diameter of

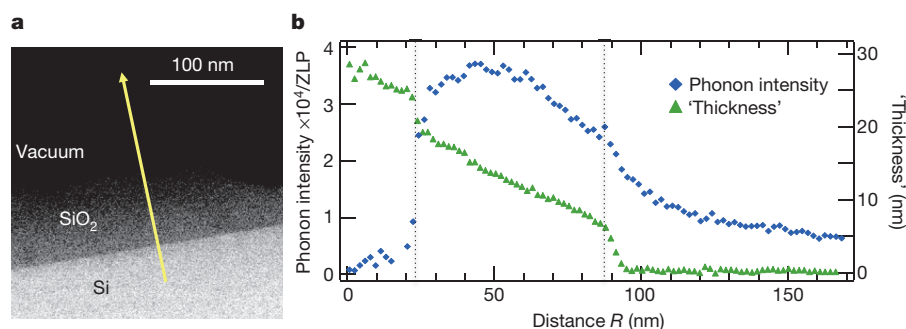


Figure 3 | Profile showing the spatial variation of the vibrational signal. **a**, HAADF image of an Si-SiO₂-vacuum sample viewed in cross-section; the yellow arrow indicates the location of the profile used for **b**. **b**, Net intensity of

the SiO₂ vibrational signal (at 138 meV) as a function of the probe position, and approximate sample thickness. The dotted vertical lines mark the locations of the Si-SiO₂ and SiO₂-vacuum interfaces.

the focused probe, the vibrational signal can be collected with the electron beam positioned outside the sample^{20–22}. This was in fact done for the titanium hydride spectrum shown in Fig. 2. The ‘aloof mode’ promises to allow radiation-sensitive materials to be analysed with no (or very limited) ionization and knock-on damage, provided that probe tails are minimized and the flux of electrons directly traversing the sample is negligible. Electrons passing tens of nanometres outside the sample can only cause low-energy excitations, and this means that radiation damage in the aloof mode can be no greater than if the sample were illuminated by infrared light. This will be especially beneficial for materials containing hydrogen, which is very prone to radiation damage due to ionization and knock-on damage by a traversing electron beam.

The spatial variation of the vibrational signal shown in Fig. 3b also includes sharp features that are much narrower than the extended tail. At the SiO₂-vacuum interface, the signal decays to about 70% of the intensity within about 10 nm from the edge, and there is a rapid decay of the vibrational signal within about 5 nm at the SiO₂-Si interface. Recent theoretical estimates suggest that spatial resolutions down to 1 Å should be attainable^{26–28} when imaging atomic vibrational states. As discussed in the Methods, the localized signal can be enhanced by selecting only higher-angle scattering events using an annular spectrometer entrance aperture. The cross-section will then be smaller than for the forward-scattered signals used in the present paper, and collecting the spectra will be more challenging experimentally, as it will require combining good energy resolution with large acceptance angles into the spectrometer. Nevertheless, the localized signal is likely to become routinely accessible with further improvements in the experimental technique, leading to a substantial increase in spatial resolution.

Very good spatial resolution should also be obtainable by using impact (Rutherford) scattering⁵, which is dominant at high scattering angles (>50 mrad), and which can readily result in energy transfers in excess of 50 meV to the nuclei of light elements, especially hydrogen. Energy-analysing the scattering²⁹ should lead to information similar to that provided by neutron vibrational spectroscopy⁴, in principle at the same ångström-scale spatial resolution as HAADF STEM³⁰, which also utilizes the Rutherford-scattered signal.

In terms of applications, the present technique promises the analysis of vibrational modes at high spatial resolution at structural features such as surfaces, interfaces, grain boundaries, nanotubes and quantum dots. It should also enable electron microscopes to directly detect hydrogen in hydrogen storage materials, polymers and biological tissues. One of the main prospects for the technique is carrying out largely damage-free analysis of radiation-sensitive materials, such as polymers or frozen hydrated biological tissues, using an aloof electron beam. The spatial resolution attained so far is about 5 nm. Future improvements are likely to include electron spectrometers with higher stability and better-corrected optics, which should improve the energy resolution of the technique to below 10 meV and the spatial resolution to below 1 nm. The potential for vibrational spectroscopy in the electron microscope thus appears to be broad and exciting.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 June; accepted 18 September 2014.

1. Stuart, B. *Infrared Spectroscopy: Fundamentals and Applications* (Wiley, 2004).
2. Griffiths, O. R. & De Haseth, J. A. *Fourier Transform Infrared Spectrometry* 2nd edn (Wiley, 2007).
3. McCreery, R. L. *Raman Spectroscopy for Chemical Analysis* (Wiley, 2000).
4. Mitchell, P. C. H., Parker, S. F., Ramirez-Cuesta, A. J. & Tomkinson, J. *Vibrational Spectroscopy With Neutrons* (World Scientific, 2005).
5. Ibach, H. & Mills, D. L. *Electron Energy Loss Spectroscopy and Surface Vibrations* (Academic, 1982).
6. Stipe, B. C., Rezaei, M. A. & Ho, W. Single molecule vibrational spectroscopy and microscopy. *Science* **280**, 1732–1735 (1998).
7. Dai, S. *et al.* Tunable photon polaritons in atomically thin van der Waals crystals of boron nitride. *Science* **343**, 1125–1129 (2014).
8. Haider, M., Braunshausen, G. & Schwann, E. Correction of the spherical aberration of a 200 kV TEM by means of a hexapole-corrector. *Optik* **99**, 167–179 (1995).
9. Krivanek, O. L., Dellby, N., Spence, A. J., Camps, R. A. & Brown, L. M. Aberration correction in the STEM. *Inst. Phys. Conf. Ser.* **153**, 35–40 (1997).
10. Haider, M. *et al.* Electron microscopy image enhanced. *Nature* **392**, 768–769 (1998).
11. Batson, P. E., Dellby, N. & Krivanek, O. L. Sub-Ångström resolution using aberration corrected electron optics. *Nature* **418**, 617–620 (2002).
12. Erni, R., Rossell, M. D., Kisielowski, C. & Dahmen, U. Atomic-resolution imaging with a sub-50-pm electron probe. *Phys. Rev. Lett.* **102**, 096101 (2009).
13. Krivanek, O. L. *et al.* Atom-by-atom structural and chemical analysis by annular dark field electron microscopy. *Nature* **464**, 571–574 (2010).
14. Suenaga, K. & Koshino, M. Atom-by-atom spectroscopy at graphene edge. *Nature* **468**, 1088–1090 (2010).
15. Zhou, W. *et al.* Direct determination of the chemical bonding of individual impurities in graphene. *Phys. Rev. Lett.* **109**, 206803 (2012).
16. Ramasse, Q. M. *et al.* Probing the bonding and electronic structure of single atom dopants in graphene with electron energy loss spectroscopy. *Nano Lett.* **13**, 4989–4995 (2013).
17. Krivanek, O. L., Lovejoy, T. C., Dellby, N. & Carpenter, R. W. Monochromated STEM with a 30 meV-wide, atom-sized electron probe. *Microscopy* **62**, 3–21 (2013).
18. Boersch, H., Geiger, J. & Stickel, W. Interaction of 25-keV electrons with lattice vibrations in LiF. Experimental evidence for surface modes of lattice vibration. *Phys. Rev. Lett.* **17**, 379–381 (1966).
19. Terauchi, M., Tanaka, M., Tsuno, K. & Ishida, M. Development of a high energy-resolution electron energy loss spectroscopy microscope. *J. Microsc.* **194**, 203–209 (1999).
20. Walls, M. G. & Howie, A. Dielectric theory of localized energy loss spectroscopy. *Ultramicroscopy* **28**, 40–42 (1989).
21. Cohen, H. *et al.* Near-field electron energy loss spectroscopy of nanoparticles. *Phys. Rev. Lett.* **80**, 782–785 (1998).
22. Garcia de Abajo, F. J. & Howie, A. Electron spectroscopy from outside – aloof beam or near field? *Inst. Phys. Conf. Ser.* **161**, 327–330 (1999).
23. Howie, A. & Milne, R. H. Excitations at interfaces and small particles. *Ultramicroscopy* **18**, 427–4334 (1985).
24. Wang, Z. L. Valence electron excitations and plasmon oscillations in thin films, surfaces, interfaces and small particles. *Micron* **27**, 265–299 (1996).
25. Garcia de Abajo, F. J. Optical excitations in electron microscope. *Rev. Mod. Phys.* **82**, 209–275 (2010).
26. Cueva, P. & Muller, D. A. Atomic-scale optical and vibrational spectroscopy with low loss EELS. *Microsc. Microanal.* **19** (suppl. 2), 1130–1131 (2013).
27. Dwyer, C. Localization of high-energy electron scattering from atomic vibrations. *Phys. Rev. B* **89**, 054103 (2014).
28. Rez, P. Is localised infrared spectroscopy now possible in the electron microscope? *Microsc. Microanal.* **20**, 671–677 (2014).
29. Lovejoy, T. C. *et al.* Energy-filtered high-angle dark field mapping of ultra-light elements. *Microsc. Microanal.* **20** (suppl. 3), 558–559 (2014).

30. Pennycook, S. J. & Nellist, P. D. (eds) *Scanning Transmission Electron Microscopy* (Springer, 2011).

Acknowledgements We thank A. Howie and J.-C. Idrobo for discussions, W. J. Bowman, J. Bruley, J. H. Butler, V. Domnich, R. A. Haber, Y. Ikuhara, M. R. Libera, D. S. Lowry and V. Nicolosi for provision of samples, J. Mardinly for help with running the instruments, our co-workers at Nion, especially N. J. Bacon, G. J. Corbin, P. J. Cramer, Z. Dellby, R. W. Hayner, P. Hrnčirik, P. Phoungphidok, M. C. Sarahan, G. S. Skone, Z. Szilagyí and T. Yoo for help with the construction of the hardware, electronics and software for HERMES, and C. Trevor of Gatan Inc. for an instability-analysing script. We also acknowledge the use of facilities within the LeRoy Eyring Center for Solid State Science at Arizona State University. Financial support for the purchase of the microscopes was provided by National Science Foundation grants DMR MRI 0821796 (Arizona State University) and DMR MRI-R2 959905 (Rutgers University). Department of Energy grant DE-SC0004954 provided support for P.A.C. and microscopy performed at Arizona State University, and Department of Energy grant DE-SC0005132 provided support for P.E.B., M.J.L. and microscopy performed at Rutgers University. Additional support was provided by the Department of Energy (grant DE-SC0007694), the Natural Sciences

and Engineering Council of Canada, the UK Engineering and Physical Research Council (capital equipment grant EP/J021156/1), Arizona State University, Rutgers University and Nion Co.

Author Contributions P.A.C., O.L.K. and P.R. initiated the project, P.A.C., O.L.K. and J.Z. prepared samples, T.A., P.E.B., N.D., O.L.K., M.J.L., T.C.L. and J.Z. obtained electron microscope spectra and images, P.A.C. and E.S. obtained infrared and Raman spectra, T.C.L., N.D., R.F.E. and J.Z. analysed EELS results, P.R. performed theoretical simulations, P.E.B., R.W.C., N.D., R.F.E., O.L.K., T.C.L. and P.R. advised on theoretical interpretation, P.E.B., O.L.K., T.C.L. and E.S. prepared the figures, and O.L.K. and P.R. wrote the paper. All the authors read and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to O.L.K. (krivanek@nion.com) or P.A.C. (crozier@asu.edu).

METHODS

Materials and sample preparation. Samples were prepared by crushing and dispersing on holey carbon grids (h-BN, SiC, TiH₂), tripod polishing (Si–SiO₂ cross-section) and ultramicrotoming (epoxy resin). In greater detail as follows.

h-BN: As described in ref. 13.

SiC: Fractured powder of mixed SiC polytype. The examined areas had thicknesses of $\sim 0.7 \times$ (mean free path of EELS integrated between 5 and 200 eV).

Si–SiO₂–vacuum interface sample: Semiconductor device test structure that was cross-sectioned and thinned using tripod polishing.

TiH₂: TiH₂ powder was crushed between two glass microscope slides and dispersed over a holey carbon film supported on a Cu grid. Smaller particles often failed to show any hydrogen signal, and may have lost their hydrogen during sample preparation. Larger particles (greater than about 0.5 μm in size) showed the hydrogen vibrational signal reliably, and we therefore typically collected spectra from thin areas of larger particles.

Epoxy resin: The sample was a microtomed section of a standard embedding resin used for TEM applications. The resin, obtained from Electron Microscopy Services (catalogue no. 14300), was a low-viscosity epoxy based on a modification of the Spurr resin³¹. The resin was microtomed into sections approximately 100 nm in thickness and supported over a holey carbon film.

Spectrum and image acquisition. Spectra in Figs 1b, 2a, 2b, 2d, 2e and 3 were obtained at Arizona State University with Nion high-energy resolution monochromated EELS systems (HERMES) using a Gatan Enfium spectrometer; spectra in Figs 1a and 2c were obtained with a HERMES system at Rutgers University using a home-modified spectrometer. Both systems were operated at 60 kV primary voltage, with a beam current incident on the sample of ~ 10 pA (and a beam current incident on the monochromator slit of ~ 500 pA), probe convergence semi-angles up to 30 mrad, and EELS collection semi-angles of typically 15 mrad.

The detailed spectrum acquisition times and operating modes used were as listed below. For a more detailed explanation of the HERMES system and its operating modes, see the Methods sections ‘Monochromator design’ and ‘Operating modes’. Figure 1a (monochromated): single acquisition of 10 ms, SR mode with $\alpha = 30$ mrad; Fig. 1b (BN peak): 6 acquisitions of 1 s each, displaced in position to randomize detector pixel variations, and later aligned in energy and summed, SR mode with $\alpha = 30$ mrad; Fig. 2a: single acquisition of 5 s, SR mode with $\alpha = 12$ mrad; Fig. 2b: single acquisition of 15 s, SR mode with $\alpha = 30$ mrad; Fig. 2c: 300 acquisitions of 0.15 s each, displaced in position to randomize detector pixel variations, and later aligned in energy and summed, SR mode with $\alpha = 16$ mrad; Fig. 2d: single acquisition of 10 s, SR mode with $\alpha = 12$ mrad; Fig. 2e: single acquisition of 50 s, AR mode, diameter of the probed sample area ~ 400 nm, acceptance half-angle ~ 50 μrad ; Fig. 3 and Extended Data Fig. 3: single acquisition of 10 s per probe position, SR mode with $\alpha = 12$ mrad.

All spectra were acquired with an energy dispersion of 1 meV per channel and with a 1 mm diameter EELS entrance aperture. The large dispersion helped to minimize the tail of the ZLP due to sideways spreading of the strong ZLP signal in the scintillator and the fibre optic coupling to the CCDs that detected the spectrum in the spectrometers used here. The images shown in Fig. 3 and Extended Data Fig. 3 were acquired using an HAADF detector spanning detection half-angles of about 80–200 mrad.

Spectrum processing and quantification. Spectra shown in this paper are displayed as-acquired, with no processing except for CCD dark image subtraction and gain normalization. Special care was taken with reference dark and gain images so that so that their statistical noise would not appreciably worsen the noise characteristics of the dark-subtracted and gain-normalized spectra.

The intensities shown in the spatial profiles (Fig. 3 and Extended Data Fig. 3) were extracted by fitting an $a\Delta E^{-r}$ function³² to the spectrum background next to the peak to be quantified (where ΔE is the energy loss and a and r are adjustable parameters), over a broad energy interval in front of the peak and a narrow interval after the peak. The quality of the fit was checked by examining the background next to the extracted peak and making sure that its intensity was negligible compared to the peak.

The approximate sample ‘thickness’ t shown in Fig. 3b and Extended Data Fig. 3b was determined by the usual method³² of computing the thickness from the ratio of the complete EELS intensity I_{total} to the intensity of the ZLP I_0 , as

$$t = \text{IMFP} \times \ln(I_{\text{total}}/I_0)$$

where IMFP is the mean free path for inelastic scattering in the examined material. The energy window used for evaluating I_{total} extended up to 250 eV for Fig. 3, and up to 190 eV for Extended Data Fig. 3. IMFP for 60 kV electrons in SiO₂ was computed using the program IMFP.m (ref. 32, appendix B) as 70 nm, IMFP in h-BN as 75 nm. The same IMFP was used for Si as for SiO₂, and the abrupt increase in the

approximate thickness seen in Fig. 3b on going from SiO₂ into Si was probably due to IMFP being smaller in Si than in SiO₂.

Monochromator design. Without a monochromator, the attainable energy resolution of EELS carried out in an electron microscope would be limited by the energy width of the electron source to about 0.25 eV for cold field emission (CFE) and 0.5 eV for Schottky sources.

Monochromators employed in electron microscopy have typically been located inside the electron gun, before the electron accelerator. Such an arrangement causes the energy of the electron beam to change whenever instabilities change the high voltage supplied to the electron gun. The end result is a deterioration of the EELS energy resolution, especially for longer acquisition times³³. The best energy resolution has therefore been reached with systems that compensate for the instabilities by decelerating the electrons before energy analysis, using a high voltage power supply that is shared by the electron gun accelerator and the spectrometer decelerator, so that changes in the high voltage of the gun are precisely compensated in the spectrometer. This type of approach has produced an energy resolution of about 3 meV in a scattering apparatus¹⁸ operating at 30 kV, and 12 meV in a substantially modified electron microscope¹⁹ operating at 60 kV. However, neither of these instruments achieved nm-level spatial resolution, their limits being about 30 μm and 100 nm (at the highest energy resolution), respectively.

The ground-potential monochromator¹⁷ used here employs a stabilization scheme that avoids decelerating the beam but manages to reach comparable energy resolution. It projects an energy-dispersed beam onto an energy-selecting slit using magnetic prisms and multipole optics consisting of about 120 independent optical elements. Because it comes after the accelerator of the electron microscope, the energy it selects does not change when the high voltage supplied to the electron gun changes. The EELS data are acquired using a magnetic prism spectrometer^{34,35} that also employs multipole optics. The system’s principal stabilization scheme consists of using a shared current supply for powering the magnetic prisms of the monochromator and of the spectrometer, with the principal coils of all the prisms connected in series. This ensures that if the current in the prism coils changes, the EEL spectrum dispersed on the final detector does not move^{17,36}. An auxiliary stabilization scheme senses the intensity of the energy-dispersed electron beam incident on the two halves of the energy-selecting slit of the monochromator, and uses the difference signal to adjust the high voltage of the microscope so that the beam remains precisely centred on the slit. The entire system has achieved 3 meV root mean square (r.m.s.) stability at a primary energy of 60 keV, which amounts to a stability of 5 parts in 10⁸. This is about a threefold improvement on the stability we were reaching a year ago (Extended Data Fig. 1), and further improvements are expected in the future. The system is also able to achieve a probe size¹⁷ of about 1 Å.

Operating modes. For spectrum acquisition, the STEM–EELS system was operated in two different modes:

(1) SR (spatially resolved) mode, in which there was a focused probe at or near the sample. The probe half-angle α was adjustable (by exciting the condenser lenses differently³⁷), and we typically used half-angles from 12 to 30 mrad. To maximize the current in the probe, the current incident on the monochromator slit was typically set to around 500 pA, which was about $2.5 \times$ the coherent current³⁸ I_c of the cold field emission (CFE) source used here. This meant that the minimum attainable probe diameter was about $2 \times$ larger than the diffraction-limited probe diameter d_d ($d_d = 0.61 \lambda/\alpha$, where λ is the electron wavelength). At 60 kV ($\lambda = 4.86$ pm) the $\alpha = 12$ mrad probe therefore had a minimum diameter of 0.5 nm, and the $\alpha = 30$ mrad probe had a minimum diameter of 0.2 nm.

The SR mode is the standard mode of operating a modern STEM: there is a small probe on the sample, and a convergent beam diffraction pattern of adjustable camera length is projected onto the high-angle annular dark field (HAADF) detector and the EELS entrance aperture. The ‘EELS entrance object’, that is, the beam crossover that is imaged and energy-dispersed by the EELS optics onto the EELS CCD detector, is formed in this mode by the last projector lens of the post-sample optical column³⁷, P4. If the crossover is too large (or has a pronounced spatially extended tail), it worsens the EELS energy resolution (or adds an energy tail to the ZLP). Therefore, it is useful to keep track of how this crossover is formed, and how the post-sample column aberrations affect its spatial extent. We typically adjusted the post-sample column to give a camera length at the EELS entrance aperture of 33 mm, which meant that a 1 mm diameter EELS aperture accepted half-angles up to 15 mrad, and that the EELS entrance object in the back-focal plane of P4 was an image of the sample-level probe magnified about $15 \times$. For 10 meV energy resolution, entrance crossover diameters of up to about 20 nm could be tolerated, and the above set-up was safely within this limit.

For samples that were moderately sensitive to radiation damage, we typically defocused the sample-level probe to about 2 nm diameter. Much wider areas of the sample could be probed in the SR mode by setting up post-sample descanning such that the beam entering the spectrometer was stationary while the beam at the sample was scanning over large areas.

(2) AR (angle-resolved) mode, in which a large area of the sample was illuminated, and the angular resolution was typically about $300\times$ higher than in the SR mode. The objective lens excitation was kept the same as in the SR mode, and the sample was moved about $16\text{ }\mu\text{m}$ below (upstream of) the objective lens crossover, resulting in about a 400 nm diameter sample area being illuminated when a 12 mrad illumination half-angle was used, and a diffraction pattern of $16\text{ }\mu\text{m}$ camera length being formed in the crossover plane. The projector lenses were adjusted to project a highly magnified ($\sim 500,000\times$) version of this pattern into the entrance aperture plane of the EEL spectrometer, giving a camera length of about 8 m . With the 1 mm EELS entrance aperture that was typically used, this corresponded to an EELS collection half-angle of about $60\text{ }\mu\text{rad}$. The P4 crossover served once more as the EELS object point. It contained an image of the illuminated sample area, demagnified about $20\times$. The illuminated area of the sample and the camera length could be varied very flexibly in this mode, simply by moving the sample up/down using the nm-precision z -drive of the microscope sample stage. The illumination half-angle of the focused beam inside the objective lens could be varied too, between about 1 and 40 mrad , and this provided additional flexibility. The magnification of the objective lens crossover projected into the EELS aperture plane was always adjusted to illuminate the whole 1 mm aperture. This ensured that the P4 crossover (whose size is inversely proportional to the size of the object projected into the EELS entrance aperture plane) was sufficiently small, so that the size of its image appearing on the spectrometer CCD would not adversely affect the attainable energy resolution.

The AR mode demagnifies a sample level-object to form the EELS entrance object. Haloes formed around the object due to aberrations of the post-sample column or stray scattering from aperture edges, and so on, are demagnified too. As a result, the AR mode is typically able to reduce the tail of the ZLP. It is, however, sensitive to sample charging. The beam is spread on the sample in the AR mode, and sample charging causes a spurious lens effect, which usually varies in time. The spurious lens can distort the shape of the ZLP, causing a loss of energy resolution, as well as an enhancement in the strength of the ZLP tail.

The peak widths of the observed optical phonon features were typically larger than the width of the ZLP. There were several reasons for this, including: (1) integrating over a range of scattering angles and hence over different energy phonons when the phonon energy varied with angle (that is, in the presence of angular dispersion), (2) the finite width of the optical phonon features, and (3) sample charging affecting the resolution in the AR mode as described above. As an example of the second effect, the FWHM of the dominant SiO_2 optical phonon in infrared spectra³⁹ is about 20 meV , that is, almost the same as the one recorded here.

Infrared, Raman and HREEL spectra of examined materials. Infrared, Raman, high resolution EELS (HREELS) or neutron vibrational spectra are available in the literature for all the materials studied here, with the exception of the epoxy resin. HREELS studies of h-BN using low-energy electrons have identified longitudinal optical (LO), shear (SH) and transverse optical (TO) phonon modes⁴⁰, and shown that the dominant LO mode energy varies with the collection angle. For the present collection geometry which emphasized low-angle scattering, the expected LO phonon energy is about 175 meV . SiO_2 has a dominant infrared peak³⁹ at $1,100\text{ cm}^{-1} = 136\text{ meV}$ (to convert cm^{-1} to meV , divide by 8.066), SiC shows a strong infrared absorption peak⁴¹ at $790\text{ cm}^{-1} = 98\text{ meV}$, and TiH_2 gives a dominant peak in neutron vibrational spectra⁴² at 150 meV . Except for the SiC spectrum, these all agree with the energies observed in this study to within 3 meV . The SiC energy discrepancy is 8 meV and is probably due to energy differences between different SiC polytypes.

A Raman spectrum from the epoxy resin is shown in Extended Data Fig. 2. It was acquired with an FTIR/FT-Raman spectrometer (Bruker IFS 66V/S) that used a $1,064\text{ nm}$ laser, and the scan covered the range $100\text{--}3,500\text{ cm}^{-1}$. The major peak is centred on $2,920\text{ cm}^{-1} = 362\text{ meV}$, in good agreement with the observed energy of the EELS peak (360 meV).

Theoretical modelling and comparison with experiment. The theory for excitation of localized vibrations by medium energy ($30\text{--}300\text{ keV}$) electrons has been given by Dwyer²⁷ and Rez²⁸. The cross-section for the vibrational scattering of $60\text{--}100\text{ keV}$ electrons is in the range of $10^3\text{--}10^4$ barns per atom pair²⁸, that is, comparable to the integrated cross-sections of EELS inner shell edges, such as the carbon K edge at 284 eV .

The spatial dependence of the signal is well described by the dielectric response to an electron beam moving parallel to a planar interface^{23–25}. The scattering probability per unit path length z for an energy loss ΔE is

$$\frac{dP(R, \omega)}{d\Delta E dz} = \frac{e^2}{2\pi^2 \epsilon_0 \hbar^2 v^2} \left\{ \text{Im} \left(-\frac{1}{\epsilon_1(\omega)} \right) \left[\ln \left(\frac{q_c v}{\omega} \right) \right] + K_0 \left(\frac{2\omega R}{v\gamma} \right) \left[\text{Im} \left(-\frac{2}{\epsilon_1(\omega) + \epsilon_2(\omega)} \right) - \text{Im} \left(-\frac{1}{\epsilon_1(\omega)} \right) \right] \right\}$$

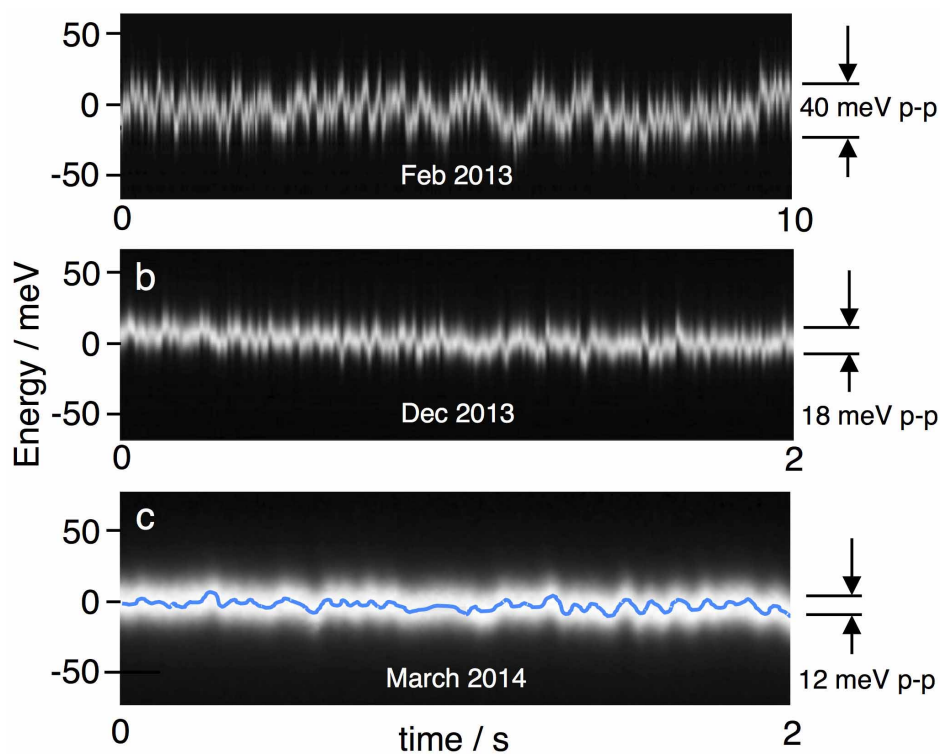
where R is the distance of the beam from medium 2, ω is the vibration frequency, q_c is a cut-off wave vector, v is the electron velocity in medium 1, $\epsilon_1(\omega)$ and $\epsilon_2(\omega)$ are the dielectric functions of media 1 and 2 respectively, and γ is the Lorentz factor. For the aloof beam, medium 1 is a vacuum and $\epsilon_1(\omega) = 1$. The behaviour with distance from the interface is then given simply by the modified Bessel function of the second kind, K_0 , as described in the main text.

A K_0 model for the drop-off of the phonon signal with distance R in the vacuum is shown in Extended Data Fig. 3, superimposed on an experimentally measured drop-off in the intensity of the LO optical phonon in h-BN beyond the edge of a BN particle (normalized to the ZLP intensity). The particle thickness, measured by medium-loss EELS (covering an energy interval $5\text{--}190\text{ eV}$), is also shown. The drop-off is probed out to a distance of 300 nm beyond the particle's edge, and the agreement with the K_0 function is very good. Because the thickness of the sample was not known accurately, the vertical scaling of the K_0 model was adjusted for the best fit. This did not change the predicted relative rate of the signal's decrease with distance.

The sampling interval for the experimental dependence was 20 nm , that is, $10\times$ coarser than for the SiO_2 -vacuum interface probed in Fig. 3, and too coarse to fully capture the spatial variation right at the particle's edge. At $>100\text{ nm}$ from the sample edge, the phonon tail follows a simple exponential decay, as shown in Extended Data Fig. 3. Closer to the edge of the sample, however, both the observed phonon signal and the K_0 function rise to much higher values than an exponential. K_0 in fact rises to infinity at $R = 0$, but the singularity is removed in practice by the finite width of the electron probe and by other effects such as high-angle scattering events not being accepted into the spectrometer. The excess signal relative to the exponential corresponds to a 'localized' vibrational signal, and a key question for the future is: how precisely can the localized signal be recorded in the presence of the 'extended' signal?

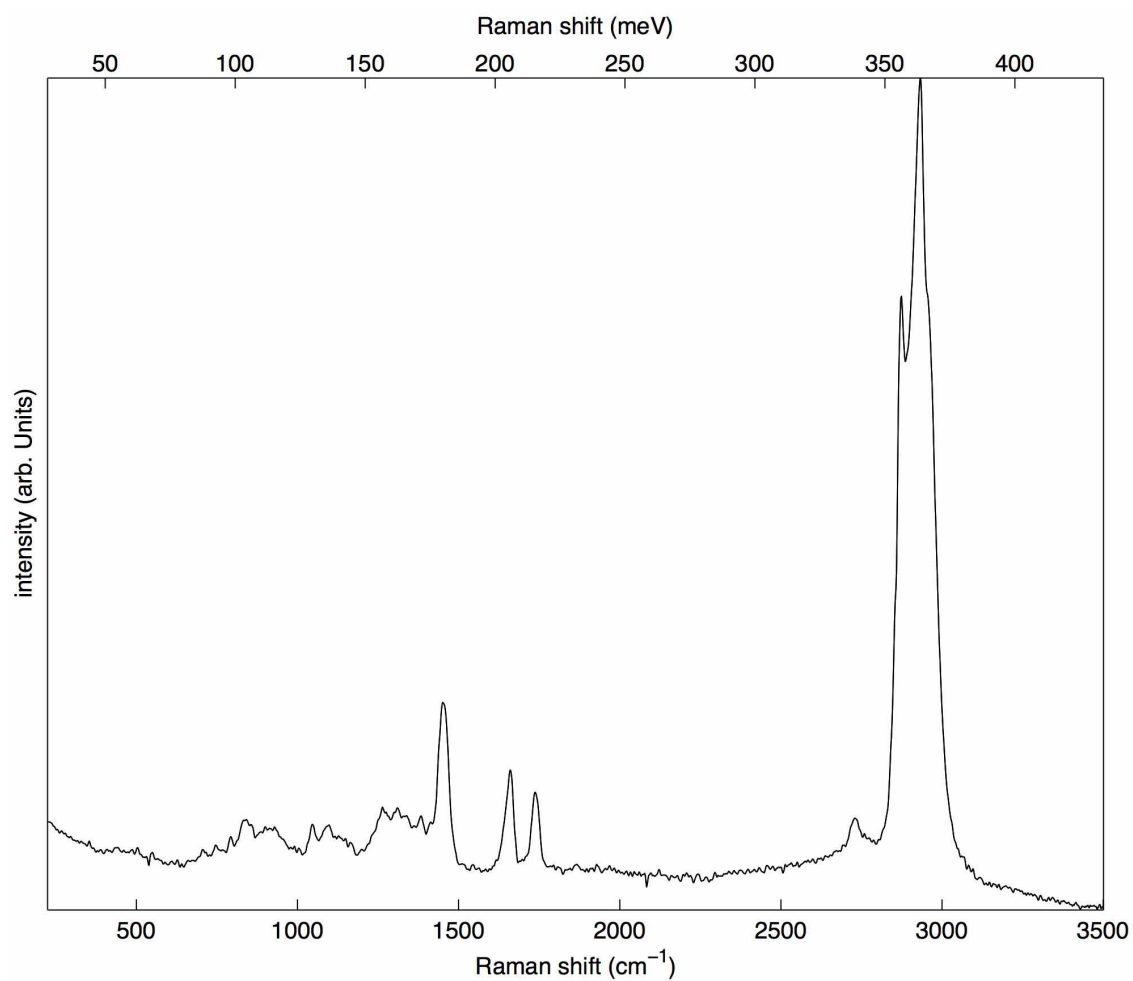
Accepting only scattering angles greater than a few mrad into the EEL spectrometer, for instance by an annular pre-EELS aperture that blocks the STEM bright field disk, should in principle suppress the extended signal and make the localized signal more visible. Collecting the localized signal efficiently will then require increasing the range of the scattering angles accepted into the spectrometer, while keeping the energy resolution the same (or improving it). This will place higher demands on the quality of the aberration correction in the post-sample part of the optical column and in the spectrometer. Such improvements are expected in the future, especially now that it has become clear that they are likely to lead to a significant increase in the spatial resolution of vibrational spectroscopy in the electron microscope.

- Spurr, A. R. A low-viscosity epoxy resin embedding medium for electron microscopy. *J. Ultrastruct. Res.* **26**, 31–43 (1969).
- Egerton, R. F. *Electron Energy Loss Spectroscopy in the Electron Microscope* 3rd edn (Springer, 2011); <http://www.tem-eels.ca/computer-programs/index.html>.
- Tiemeijer, P. C., van Lin, J. H. A., Freitag, B. H. & de Jong, A. F. Monochromized 200 kV (S)TEM. *Microsc. Microanal.* **8** (suppl. 2), 70–71 (2012).
- Krivanek, O. L., Gubbens, A. J., Dellby, N. & Meyer, C. E. Design and first applications of a post-column imaging filter. *Microsc. Microanal. Microstruct.* **3**, 187–199 (1992).
- Gubbens, A. J. *et al.* The GIF Quantum, a next generation post-column imaging energy filter. *Ultramicroscopy* **110**, 962–970 (2010).
- Krivanek, O. L. *et al.* Towards sub-10 meV energy resolution STEM-EELS. *Inst. Phys. Conf. Ser.* **522**, 012023 (2014).
- Krivanek, O. L. *et al.* An electron microscope for the aberration-corrected era. *Ultramicroscopy* **108**, 179–195 (2008).
- Krivanek, O. L., Chisholm, M. F., Dellby, N. & Murfitt, M. F. in *Scanning Transmission Electron Microscopy: Imaging and Analysis* (eds Pennycook, S. J. & Nellist, P. D.) 613–656 (Springer, 2011).
- Bock, J. & Gouq-Jen, S. Interpretation of the infrared spectra of fused silica. *J. Am. Ceram. Soc.* **53**, 69–73 (1970).
- Oshima, C. & Nagashima, A. Ultra-thin epitaxial films of graphite and hexagonal boron nitride on solid surfaces. *J. Phys. Condens. Matter* **9**, 1–20 (1997).
- Li, P. J. *et al.* Structural characterization of nm SiC films grown on Si. *Appl. Phys. Lett.* **62**, 3135–3137 (1993).
- Bashkin, I. O. *et al.* Hydrogen interaction and bound multiphonon states in vibrational spectra of titanium hydrides. *Z. Phys. Chem.* **179**, 335–342 (1993).

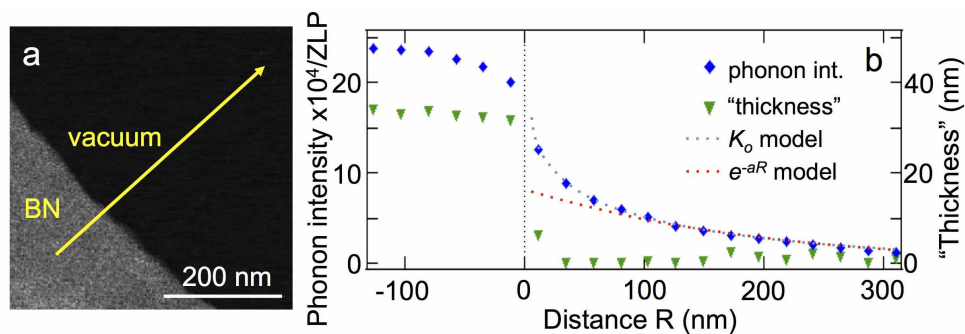


Extended Data Figure 1 | Analysis of total system instabilities, performed by plotting the position of the ZLP on the EEL spectrometer detector as a function of time. **a**, Using the Gatan Enfium EELS and original power supplies: ~ 50 meV peak-to-peak and 12 meV r.m.s. instability. **b**, As **a** but with improved-stability power supplies: ~ 18 meV peak-to-peak and 4.5 meV r.m.s.

instability. **c**, Single magnetic sector EELS with Nion power supplies: ~ 12 meV peak-to-peak and 3 meV r.m.s. instability. Panels **a** and **b** show the position of the ZLP on the EELS detector as a function of time, marking each position of the ZLP with a single point; **c** shows the whole ZLP profile, together with a blue trace that follows the instantaneous position of the centre of the profile.



Extended Data Figure 2 | Raman spectrum of epoxy resin. The major peak is centred on $2,920\text{ cm}^{-1} = 362\text{ meV}$.



Extended Data Figure 3 | Fall-off of the h-BN LO phonon signal in the vacuum, probed out to 300 nm. **a**, HAADF image of an edge of a BN particle of nearly constant thickness, with the probed line indicated by a yellow arrow. **b**, Shown are the intensity of the BN phonon signal peak (at 173 meV)

along the probed line (blue diamonds), the K_0 model of the phonon signal decay in vacuum (dotted blue line), the exponential model of the same (dotted red line), and approximate sample thickness (green triangles). The edge of the particle is at $R = 0$.

Global covariation of carbon turnover times with climate in terrestrial ecosystems

Nuno Carvalhais^{1,2}, Matthias Forkel¹, Myroslava Khomik^{1,3}, Jessica Bellarby^{4,5}, Martin Jung¹, Mirco Migliavacca^{1,6}, Mingquan Mu⁷, Sassan Saatchi⁸, Maurizio Santoro⁹, Martin Thurner¹, Ulrich Weber¹, Bernhard Ahrens¹, Christian Beer^{1,10}, Alessandro Cescatti¹¹, James T. Randerson⁷ & Markus Reichstein¹

The response of the terrestrial carbon cycle to climate change is among the largest uncertainties affecting future climate change projections^{1,2}. The feedback between the terrestrial carbon cycle and climate is partly determined by changes in the turnover time of carbon in land ecosystems, which in turn is an ecosystem property that emerges from the interplay between climate, soil and vegetation type^{3–6}. Here we present a global, spatially explicit and observation-based assessment of whole-ecosystem carbon turnover times that combines new estimates of vegetation and soil organic carbon stocks and fluxes. We find that the overall mean global carbon turnover time is 23^{+7}_{-4} years (95 per cent confidence interval). On average, carbon resides in the vegetation and soil near the Equator for a shorter time than at latitudes north of 75° north (mean turnover times of 15 and 255 years, respectively). We identify a clear dependence of the turnover time on temperature, as expected from our present understanding of temperature controls on ecosystem dynamics. Surprisingly, our analysis also reveals a similarly strong association between turnover time and precipitation. Moreover, we find that the ecosystem carbon turnover times simulated by state-of-the-art coupled climate/carbon-cycle models vary widely and that numerical simulations, on average, tend to underestimate the global carbon turnover time by 36 per cent. The models show stronger spatial relationships with temperature than do observation-based estimates, but generally do not reproduce the strong relationships with precipitation and predict faster carbon turnover in many semi-arid regions. Our findings suggest that future climate/carbon-cycle feedbacks may depend more strongly on changes in the hydrological cycle than is expected at present and is considered in Earth system models.

The largest global gross exchanges of carbon occur at the interface between the atmosphere and the terrestrial biosphere⁷. Changes in the net exchange of CO₂ between the land and the atmosphere may provide positive or negative feedbacks to increasing atmospheric CO₂ concentrations and, thus, changes in climate^{1,8}. The response of the net exchange of CO₂ to climate depends on the response of carbon uptake (gross primary production (GPP)) and on how carbon residence times change simultaneously. It is thus of importance to quantify the time that carbon resides in terrestrial ecosystems and its spatial covariation with climate. Furthermore, global modelling studies show a stronger convergence of GPP estimates in comparison with wider ranges in whole ecosystem carbon stocks⁹, reflecting the strong spread in the modelled residence times of carbon.

In steady state, that is, when the exchange of carbon between two reservoirs is balanced, the turnover time (τ , yr) equals the mean residence time¹⁰. Assuming a balance between assimilation and losses of ecosystem carbon, τ can be calculated from the total reservoir size (C_{total} , kgC m⁻²) and the influx or the outflux (kgC m⁻² yr⁻¹) as¹⁰

$$\tau = C_{\text{total}}/\text{flux} \quad (1)$$

For terrestrial ecosystems, the total reservoir size equals the carbon stocks in vegetation and soils. The influx is the carbon uptake through gross primary production GPP and the outflux includes all carbon losses (terrestrial ecosystem respiration, fire emissions, lateral export and so on). We relax the strict steady-state assumption, calling τ the apparent whole ecosystem turnover time, computed as the ratio of C_{total} to GPP, and interpret the quantity as an emergent diagnostic at the ecosystem level (Methods section on turnover times). We note that the turnover time, or mean residence time, of carbon in ecosystems emerges from the turnover of compartments varying greatly in their individual turnover times^{5,11} (for example leaves, wood, different soil organic carbon fractions). Hence, carbon allocation, leaf, root and wood turnover, plant mortality and soil carbon decomposition are key processes that regulate the terrestrial turnover times, and which are controlled by climate variability¹².

Here we combine and enhance recently derived estimates of the carbon stocks in vegetation and soil to obtain global total terrestrial ecosystem carbon stocks and their observation-based uncertainties at 0.5° resolution (Methods section on global carbon estimates). We merge remote-sensing-based carbon stock estimates for tropical and Northern Hemisphere vegetation (Methods section on vegetation carbon) with enhanced soil organic carbon estimates based on the Harmonized World Soil Database (HWSD) and a dedicated circumpolar soil organic carbon map (which better accounts for carbon in permafrost-affected high-latitude soils; Methods section on soil organic carbon). Total soil organic carbon stocks are estimated down to full depth (that is, beyond the 100 cm depths often reported, but see Methods section on soil organic carbon). Total carbon stocks (soils and vegetation) amount to $2,807^{+855}_{-555}$ Pg of carbon C (PgC) and are predominantly within tropical forests, which contain almost 25% of the total global stocks, followed by boreal forests (18%) (Methods section on global carbon estimates). Per unit area, the total carbon stocks vary largely between and within biomes (Fig. 1a), where tropical forests and northern high latitudes exhibit the highest stocks. Substantial uncertainties are located in tundra regions (interquartile range over the mean, ~38%) and in tropical savannahs and grasslands (~30%) (Extended Data Fig. 1 and Extended Data Table 1).

Using these total ecosystem stocks and the observation-based GPP estimates in equation (1), we derive a global τ of 23 yr (ranging between 18 yr (percentile 2.5) and 29 yr (percentile 97.5)). We note that this duration is an estimation of the mean residence time of a carbon atom in terrestrial ecosystems from its initial fixation by photosynthesis until its respiratory (including autotrophic respiration) or non-respiratory

¹Max Planck Institute for Biogeochemistry, Hans Knöll Strasse 10, 07745 Jena, Germany. ²Departamento de Ciências e Engenharia do Ambiente, DCEA, Faculdade de Ciências e Tecnologia, FCT, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal. ³School of Geography and Earth Sciences, McMaster University, Hamilton, Ontario L8S 4K1, Canada. ⁴Institute of Biological and Environmental Sciences, School of Biological Sciences, University of Aberdeen, Aberdeen AB24 3UU, UK. ⁵Lancaster Environment Centre, Lancaster University, Bailrigg, Lancaster LA1 4YQ, UK. ⁶Remote Sensing of Environmental Dynamics Lab, DiSAT, University of Milano-Bicocca, Piazza della Scienza 1, 20126 Milan, Italy. ⁷Department of Earth System Science, University of California Irvine, Irvine, California 92697, USA. ⁸Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA. ⁹Gamma Remote Sensing, Worbstrasse 225, 3073 Gümligen, Switzerland. ¹⁰Department of Applied Environmental Science and Bolin Centre for Climate Research, Stockholm University, Svante Arrhenius väg 8, 10691 Stockholm, Sweden. ¹¹European Commission, Joint Research Centre, Institute for Environment and Sustainability, Climate Risk Management Unit, Via E. Fermi, 2749, I-21027 Ispra, Italy.

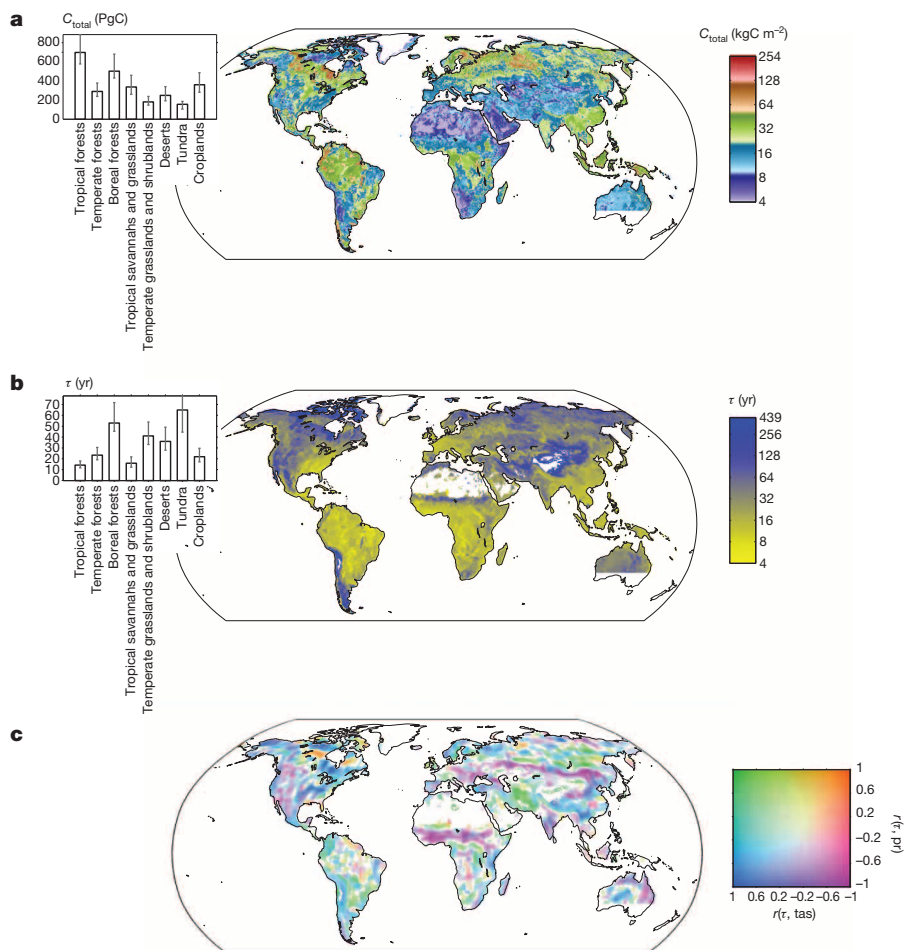


Figure 1 | Global distributions of total ecosystem carbon, turnover times of carbon in terrestrial ecosystems and its spatial covariation with climate variables. Global distribution of estimated total ecosystem carbon (C_{total}) density in each grid cell (kgC m^{-2}), and total ecosystem carbon mass per biome (PgC) (a), median of turnover times of carbon (τ) in terrestrial ecosystems (b), and the spatial covariation (r , Pearson correlation coefficient) of τ with climate variables (c). The insets in a and b show the C_{total} and τ estimates per biome according to previous classifications (Extended Data Table 1); the uncertainty bars per biome report the 95% confidence intervals (CI, ranging between percentiles 2.5 and 97.5). The ranges in C_{total} span 4 kgC m^{-2} (approximately the 5th percentile) to 254 kgC m^{-2} (maximum estimate of mean carbon stocks), and the colour code in the τ map is binned between $\tau \leq 4 \text{ yr}$ and 439 yr (the 99th percentile). Spatial correlations (moving window of 5.5° by 5.5°) with temperature (tas) and precipitation (pr) are shown only for confidence levels above 95%. The missing regions in southern Australia and New Zealand are due to missing data in the vegetation carbon data set (Methods section on total vegetation carbon).

loss. A previous collection of global estimates of net primary production, carbon in soils and carbon in vegetation allows an approximate mean estimate of τ of $21 \pm 7 \text{ yr}$, assuming 50% of autotrophic respiration costs¹³. Still, the spatial variation of τ observed in our analysis spans almost two orders of magnitude, that is, between $\sim 7 \text{ yr}$ (first percentile) and 439 yr (99th percentile) (Fig. 1b and Methods section on mean carbon turnover times).

The longest turnover times are found in cold biomes (tundra, $65^{+13}_{-20} \text{ yr}$; boreal forests, 53^{+20}_{-8} yr), in temperate grassland and shrubland (41^{+13}_{-9} yr) and in desert regions (36^{+14}_{-9} yr), whereas tropical forests and savannah exhibit the shortest turnover times (14^{+4}_{-3} and 16^{+6}_{-4} yr , respectively) (Table 1). Using localized regression analysis (Methods section on correlation analysis), we find that, spatially, τ covaries significantly ($P < 0.05$)

Table 1 | Total ecosystem turnover times of carbon for the globe, per biome

Biome type	τ (yr)		
	Mean	P 2.5	P 97.5
Tropical forests	14.2	11.6	18.2
Temperate forests	23.5	18.9	30.8
Boreal forests	53.3	45.4	73.4
Tropical savannahs and grasslands	16.0	12.2	22.1
Temperate grasslands and shrublands	41.3	32.8	54.6
Deserts	36.3	27.6	49.9
Tundra	65.2	44.7	78.0
Croplands	22.1	17.0	30.1
Wetlands	19.7	15.2	26.7
Total	22.5	18.1	29.4

Total ecosystem turnover times of carbon per biome, estimated using equation (1) and stocks and fluxes aggregated per biome. Data estimates of τ were aggregated by biomes defined previously (Extended Data Table 1), and the ranges reported are the 2.5th (P 2.5) and 97.5th (P 97.5) percentiles from the ensemble of τ estimates, which can be interpreted as the confidence intervals in these estimates. The total represents the global τ including all biomes.

with mean annual temperature or with total annual precipitation in 86% of the globe (Fig. 1c and Supplementary Information Section 2). There is, however, a strong variability in the spatial correlations between τ and temperature and τ and precipitation (Fig. 1c and Extended Data Fig. 2). Negative correlations between temperature and τ are widely observed, and can be linked to the expected decomposition responses to temperature^{14–16}. However, significant positive correlations emerge in regions of forest/herbaceous cover transitions (or patchiness) and in warm arid environments (Supplementary Information Section 4), where precipitation shows the strongest correlations with τ , indicating that moisture effects may dominate and override temperature effects. No clear dominant patterns are observed in tropical forests, suggesting that climate has a limited effect on the spatial variability of τ there and that nutrient availability¹⁷ or natural and human disturbances^{18,19}, or both, have greater effects. Globally, we observe a higher frequency of stronger spatial correlations between τ and precipitation (in $\sim 55\%$ of land grid cells) than between τ and temperature ($\sim 45\%$) (Extended Data Figs 3 and 4).

Turnover times vary considerably with latitude, ranging from 255 yr (mean τ above 75° N) in the high northern latitudes to 15 yr in the equatorial tropics (Fig. 2a). We find that the most rapid latitudinal changes exist between the sub-Arctic zones and the temperate zones, and near the tropical circles (between 20° and 40° N). Within the tropics (between 20° N and 20° S), the variations in τ are comparatively minor. In the Northern Hemisphere, in the transition zone between 50° and 65° N , the spatial covariation of τ is strongest with temperature, but south of 50° N precipitation is the dominant associated variable (until the Equator and below 40° S (Fig. 2c, d, in blue)). Across all latitudes, higher precipitation is associated with shorter residence times (negative partial correlations), whereas correlations between τ and temperature are more variable across latitudes, being low in the northern tropical zone and

also north of 60° N. Overall, across the latitudinal range, the spatial correlations of τ with temperature and precipitation are significant, but only low to moderately so. We note that carbon turnover in ecosystems will depend on the time-integrated effect of climate variables and summary statistics, and that mean total precipitation and mean temperature can serve only as simple proxies.

The significance of soil organic carbon stocks in explaining the spatial variability of τ is pervasive (Extended Data Fig. 5b). For approximately 80% of the land surface, τ covaries more strongly with soil carbon stocks than with vegetation carbon stocks. However, the residence times of carbon in terrestrial ecosystems should tend to increase with vegetation longevity, and with allocation towards woody biomass. Hence, we expected extensive positive correlations of τ with tree cover. Nevertheless, we also find negative correlations (Extended Data Fig. 5a). Precipitation, which is associated with tree cover, could overshadow the tree effect by increasing turnover times disproportionately, but the negative correlations persist even when we control for precipitation (Extended Data Fig. 6 and Supplementary Information Section 3). An increasing probability of fire related to increasing fuel loads with above-ground biomass^{20,21}, thereby reducing turnover times, is a possible explanation for this apparent paradox. Others are the contribution of trees to wetter microclimates^{22,23} and increasing nutrient availability^{23,24} in regions of low tree density and transition regions. Additionally, other factors like natural and anthropogenic disturbances²⁵ or management activities²⁶ can accelerate rates of turnover and, consequently, reduce mean residence times.

We calculated the turnover times of carbon (equation (1) in models from the Coupled Model Intercomparison Project Phase 5 (CMIP5)) (see Methods section on CMIP5 and Supplementary Information Section 5). The broad latitudinal patterns of τ in the CMIP5 ensemble (Fig. 2b) are consistent with the observations (Fig. 2a) (Pearson correlation coefficient, $r = 0.88$; $P < 0.0001$), but with a mean underestimation bias in the latitudinal profile of 47% (normalized average error). However, the zonal mean carbon turnover times vary by a factor of 2 to 40 across the analysed CMIP5 models (Fig. 2b). The models differ strongly with respect to correlations of τ with their (modelled) climate variables²⁷. Across almost all latitudes, we find a range from positive to negative correlations for both temperature and precipitation. In the Northern Hemisphere, model-derived correlations with temperature tend to be more negative than the observation-derived correlations²⁷ (Fig. 2c). For latitudes below 50° N, the model-derived correlations with precipitation are more positive when compared with the observation-derived correlations (Fig. 2d). In the tropical zones, the CMIP5 ensemble predicts increasing turnover times associated with higher precipitation, which contrasts with the observation-derived estimates.

Overall, the CMIP5 models correlate with the observation-derived estimates of τ ($r^2 = 0.38$, $P < 0.001$), but exhibit shorter turnover times (Fig. 3); this is reflected also in the global turnover times ($\sim 36\%$ lower; Extended Data Table 2). The bias is particularly pronounced in the high northern latitudes and in the seasonally dry biomes of northern tropical Africa, North America, central southern Asia and east Australia. An underestimation of turnover times in the high latitudes can potentially be

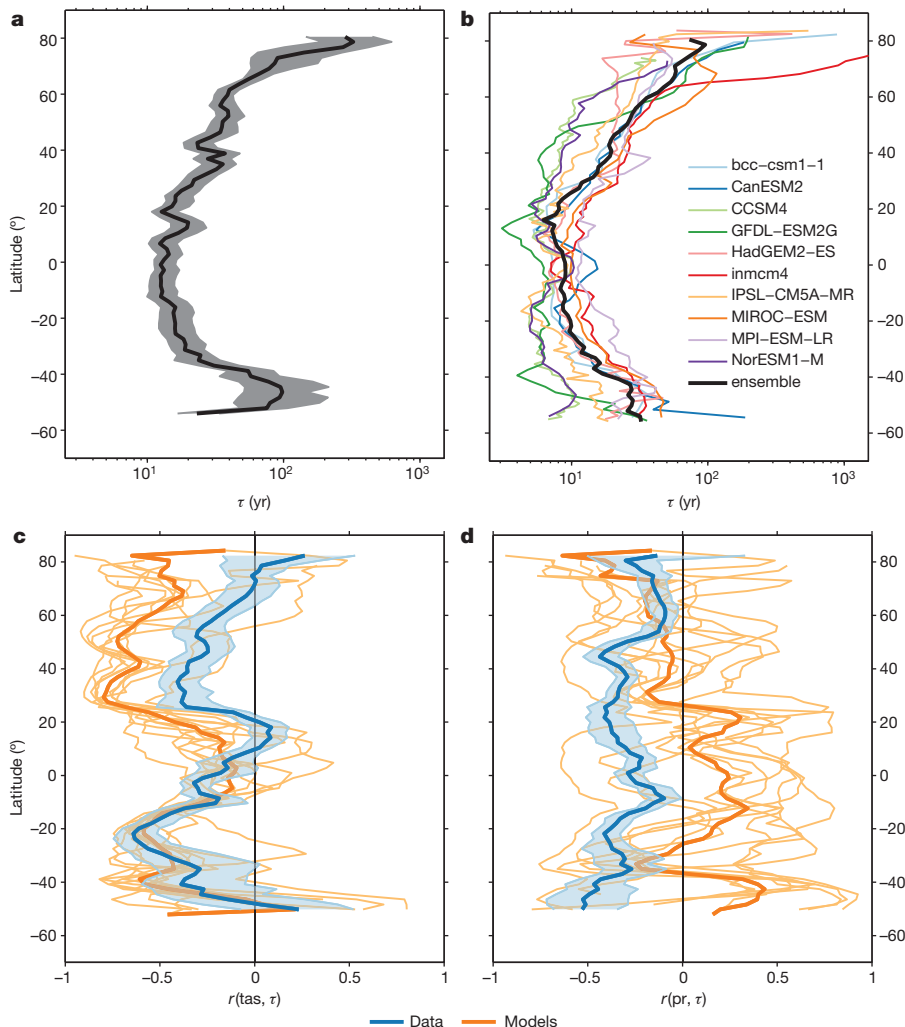


Figure 2 | Latitudinal gradients of whole ecosystem turnover times of carbon and associations to temperature and precipitation from data and models. Evaluation of latitudinal patterns and climate association of τ for data and models. The latitudinal gradients in τ from data (a) and from models (b) show distinctive associations with temperature (tas, c) and precipitation (pr, d). For consistency, the temperature and precipitation data sets considered for the model analysis are also model outputs (Methods section on CMIP5). The comparisons are based on partial correlations, controlling for precipitation when evaluating the association of τ with temperature (and vice versa), and are performed at the spatial scale of the NorESM1-M model output, to minimize artefacts in the correlations caused by differences in spatial resolution.

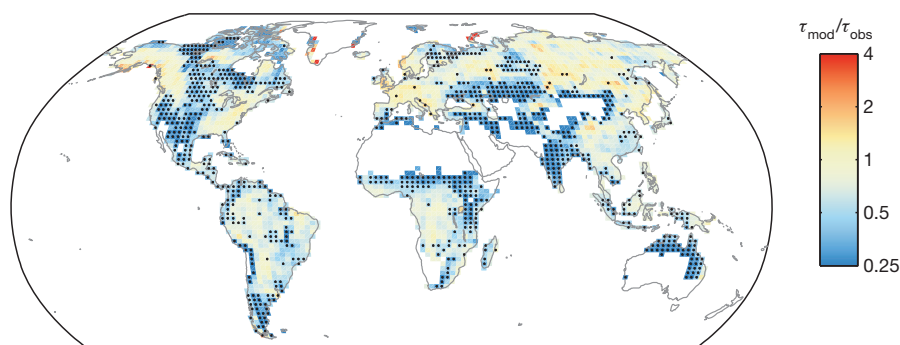


Figure 3 | Biases in whole-ecosystem carbon turnover times between models (τ_{mod}) and observation-derived (τ_{obs}) data ensembles. Stippling indicates locations where fewer than one-quarter of the models are within the 5th and 95th percentiles of the data. Here the deserts are filtered out according to a Köppen–Geiger classification (Extended Data Table 2) and to a minimum GPP of $10 \text{ gC m}^{-2} \text{ yr}^{-1}$. The missing regions in southern Australia and New Zealand are due to missing data in the vegetation carbon data set (see Methods).

explained by either neglecting, or having an incomplete representation of, permafrost processes^{28,29}, although this can only partly explain differences in biases between North America and Siberia. Globally, in 31% of the land grid (35% of the global land area analysed), fewer than one-quarter of the models are within the confidence intervals of the data. Even assuming a 50% error in the observed carbon stocks and, consequently, in turnover times, this would not explain modelled turnover times differing by more than a factor of two from observations. Furthermore, despite representing soil organic carbon pools with long residence times, CMIP5 models do not provide an explicit representation of soil organic carbon vertical profiles. This could partly explain the observed differences but not the systematic underestimation of soil organic carbon up to 1 m depth in northern latitudes (Extended Data Fig. 7), which warrants attention to the representation of soil organic pools and vertical profiles in models.

Biases in simulated climate may also lend significant biases to model estimates of τ , although a comparison of differences in τ and spatial covariations with climate reveals that models fall short in describing the climate responses seen in the observation-derived data (Supplementary Information Section 6). Other possible reasons for this pronounced model bias may include responses to, or biases in, modelled soil moisture²⁷ or insufficient sensitivity of decomposition to drought. Moreover, adaptation of vegetation to dry conditions includes leaf sclerophylly, long leaf lifespans and higher wood densities in shrubs, which together lead to increased turnover times. In addition, interactions with nutrient cycles (for example that of nitrogen) may slow the turnover of carbon in ways which are not represented in models³⁰. The spatial analysis shown here does not imply that the relationships between τ and climate factors are the same in the temporal dimension: these relationships emerge from the effects of climate—and other factors—through time, to which models should be comparable. In this regard, the emergence of appropriate model–data integration frameworks is essential for a consistent transfer of information from observation-based estimates of τ to modelling approaches³¹.

We have presented an observation-based estimate of the total terrestrial carbon pool size and whole-ecosystem carbon turnover times and its spatial variation at 0.5° resolution with associated uncertainties. Our findings suggest significant hydrological control of carbon turnover, probably as relevant as temperature, adding to the well-known coupling between carbon and water cycling for photosynthesis, and calling for a better understanding of changes with the hydrological cycle. Although the ensemble mean of state-of-the-art coupled climate/carbon-cycle models reproduces the temperature-driven latitudinal patterns of carbon turnover times, we note an important underestimation bias and differences between models of more than one order of magnitude. The pronounced underestimation of whole-ecosystem carbon turnover times in semi-arid regions calls for a more accurate description of hydrological processes and water–carbon interactions. We expect that improved representations of the adaptation of vegetation to water availability, fire dynamics, and physicochemical and microbial soil organic carbon stabilization mechanisms³², in addition to permafrost dynamics, will probably help to address the aforementioned biases. Overall, these results

emphasize the role of water on the carbon dynamics in the terrestrial biosphere and suggest that future climate/carbon-cycle feedbacks will be more sensitive to changes in the water cycle than expected and represented in state-of-the-art models.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 November 2013; accepted 30 July 2014.

Published online 24 September 2014.

1. Friedlingstein, P. *et al.* Climate-carbon cycle feedback analysis: results from the (CMIP)-M-4 model intercomparison. *J. Clim.* **19**, 3337–3353 (2006).
2. Ciais, P. *et al.* in *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. *et al.*) 465–570 (Cambridge Univ. Press, 2013).
3. King, A. W., Post, W. M. & Wullschlegel, S. D. The potential response of terrestrial carbon storage to changes in climate and atmospheric CO_2 . *Clim. Change* **35**, 199–227 (1997).
4. Sitch, S. *et al.* Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. *Glob. Change Biol.* **9**, 161–185 (2003).
5. Trumbore, S. Age of soil organic matter and soil respiration: radiocarbon constraints on belowground C dynamics. *Ecol. Appl.* **10**, 399–411 (2000).
6. Friend, A. D. *et al.* Carbon residence time dominates uncertainty in terrestrial vegetation responses to future climate and atmospheric CO_2 . *Proc. Natl Acad. Sci. USA* **111**, 3280–3285 (2014).
7. Denman, K. L. *et al.* in *Climate Change 2007: The Physical Science Basis* (eds Solomon, S. *et al.*) 499–587 (Cambridge Univ. Press, 2007).
8. Heimann, M. & Reichstein, M. Terrestrial ecosystem carbon dynamics and climate feedbacks. *Nature* **451**, 289–292 (2008).
9. Anav, A. *et al.* Evaluating the land and ocean components of the global carbon cycle in the CMIP5 earth system models. *J. Clim.* **26**, 6801–6843 (2013).
10. Rodhe, H. in *Global Biogeochemical Cycles* (eds Charlson, R. J., Butcher, S. S., Orians, G. H. & Wolfe, G. V.) Ch. 4 (Academic, 1992).
11. Malhi, Y., Saatchi, S., Girardin, C. & Aragão, L. E. O. C. in *Amazonia and Global Change* (eds Keller, M., Bustamante, M., Gash, J. & Silva Dias, P.) 355–372 (American Geophysical Union, 2009).
12. Trumbore, S. Carbon respired by terrestrial ecosystems — recent progress and challenges. *Glob. Change Biol.* **12**, 141–153 (2006).
13. Sundquist, E. T. in *The Carbon Cycle and Atmospheric CO_2 : Natural Variations, Archean to Present* (eds Sundquist, E. T. & Broecker, W. S.) 5–59 (American Geophysical Union, 1985).
14. Kätterer, T., Reichstein, M., Andren, O. & Lomander, A. Temperature dependence of organic matter decomposition: a critical review using literature data analyzed with different models. *Biol. Fertil. Soils* **27**, 258–262 (1998).
15. Davidson, E. A. & Janssens, I. A. Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. *Nature* **440**, 165–173 (2006).
16. Bond-Lamberty, B. & Thomson, A. Temperature-associated increases in the global soil respiration record. *Nature* **464**, 579–582 (2010).
17. Cleveland, C. C. & Townsend, A. R. Nutrient additions to a tropical rain forest drive substantial soil carbon dioxide losses to the atmosphere. *Proc. Natl Acad. Sci. USA* **103**, 10316–10321 (2006).
18. Houghton, R. A. Revised estimates of the annual net flux of carbon to the atmosphere from changes in land use and land management 1850–2000. *Tellus B* **55**, 378–390 (2003).
19. Nepstad, D. C. *et al.* Large-scale impoverishment of Amazonian forests by logging and fire. *Nature* **398**, 505–508 (1999).
20. Thonicke, K., Venevsky, S., Sitch, S. & Cramer, W. The role of fire disturbance for global vegetation dynamics: coupling fire into a dynamic global vegetation model. *Glob. Ecol. Biogeogr.* **10**, 661–677 (2001).
21. Krawchuk, M. A. & Moritz, M. A. Constraints on global fire activity vary across a resource gradient. *Ecology* **92**, 121–132 (2011).
22. Vetaas, O. R. Micro-site effects of trees and shrubs in dry savannas. *J. Veg. Sci.* **3**, 337–344 (1992).

23. Joffre, R. & Rambal, S. How tree cover influences the water-balance of Mediterranean rangelands. *Ecology* **74**, 570–582 (1993).
24. Belsky, A. J. Influences of trees on savanna productivity—tests of shade, nutrients, and tree-grass competition. *Ecology* **75**, 922–932 (1994).
25. Fahey, T. J. *et al.* The biogeochemistry of carbon at Hubbard Brook. *Biogeochemistry* **75**, 109–176 (2005).
26. Bondeau, A. *et al.* Modelling the role of agriculture for the 20th century global terrestrial carbon balance. *Glob. Change Biol.* **13**, 679–706 (2007).
27. Todd-Brown, K. E. O. *et al.* Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations. *Biogeosciences* **10**, 1717–1736 (2013).
28. Vonk, J. E. & Gustafsson, O. Permafrost-carbon complexities. *Nature Geosci.* **6**, 675–676 (2013).
29. Koven, C. *et al.* On the formation of high-latitude soil carbon stocks: Effects of cryoturbation and insulation by organic matter in a land surface model. *Geophys. Res. Lett.* **36**, L21501 (2009).
30. Janssens, I. A. *et al.* Reduction of forest soil respiration in response to nitrogen deposition. *Nature Geosci.* **3**, 315–322 (2010).
31. Xia, J. Y., Luo, Y. Q., Wang, Y. P. & Hararuk, O. Traceable components of terrestrial carbon storage capacity in biogeochemical models. *Glob. Change Biol.* **19**, 2104–2116 (2013).
32. Wieder, W. R., Bonan, G. B. & Allison, S. D. Global soil carbon projections are improved by modelling microbial processes. *Nature Clim. Change* **3**, 909–912 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank C. Jones for comments that improved the manuscript. We are grateful to A. Ito, D. Zaks and S. Del Grosso for sharing their NPP data sets with us. We thank S. Schott for figure editing. We acknowledge support by the European Union (FP7) through the projects GEOCARBON (283080), CARBONES (242316) and EMBRACE (283201) and an ERC starting grant QUASOM (ERC-2007-StG-208516).

Author Contributions N.C. and M.R. designed the study and are responsible for the integrity of the work as a whole. N.C., M.F. and M. Migliavacca performed analysis and calculations. N.C. and M.R. mainly wrote the manuscript. M.K. and J.B. contributed to interpreting and processing the soil databases. M.T., M.S. and S.S. contributed to the vegetation carbon stocks datasets and interpretation. M.J. contributed to the GPP datasets and interpretation. C.B., M. Mu, M.T. and U.W. contributed to data provision, analysis or data processing. A.C., B.A., M.F., M.J. and J.T.R. contributed to analysis design and interpretation. All authors discussed and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.C. (nuno.carvalhais@bgc-jena.mpg.de).

METHODS

Estimates of total soil organic carbon based on global databases. The Harmonized World Soil Database³³ condenses a comprehensive collection of geographic information on soil physical and chemical properties from regional and national inventories all over the world. The HWSD is organized in mapping units, each consisting of particular combinations of different soils referred to as 'soil ID' from here on. For every soil ID, among other variables, the database reports texture, bulk density and concentration of organic carbon for the top (0 to 10 or 30 cm) and subsoil (from 30 cm to 1 m depth) layers. Estimates of total organic carbon for each soil ID can then be computed per layer as follows:

$$\text{SOC} = \frac{\text{OC}}{100} D \left(1 - \frac{G}{100} \right) \text{BD}$$

Here soil organic carbon stocks (SOC, kg C m^{-2}) is estimated from organic carbon content (OC, wt%), layer thickness (D , m), gravel content (G , vol%) and bulk density (BD, kg m^{-3}). Such an approach allows for estimates of SOC in the top layer (0–30 cm) and in the subsoil layer (30–100 cm). We used these two estimates to fit two empirical models of cumulative SOC (equations (2) and (3)), which were then integrated until the full soil depth (D_f) per soil ID was reached, as follows:

$$\log(\text{SOC}) = K \log(D) + I \quad (2)$$

$$\text{SOC} = a \log(Db + 1) \quad (3)$$

Here K , I , a and b are empirical parameters estimated per soil ID. Estimates using equation (2) follow ref. 34 for the model with the least mean predictive error (see table 1 in ref. 34). Additionally, on the basis of *in situ* observations of SOC (ref. 35), we included an alternative model formulation that has shown a strong fit to data, as well as a faster saturation of cumulative SOC with depth (Supplementary Fig. 1). Here the latter model represents a more conservative estimation of full-depth SOC.

The full depth of the soil (D_f) was extracted from the Global Soil Texture And Derived Water-Holding Capacities database³⁶. This database contains standardized values of soil depth and textures for the globe, which were selected for the same soil types in the same continents, according to continents defined in ref. 37.

The HWSD was still a work in progress at the time of our study, such that data from certain regions in the world still needed updating and were therefore considered less reliable. Two such regions were North America and northern Eurasia³³. Therefore, as an alternative to the HWSD, we also considered the Northern Circumpolar Soil Carbon Database^{38,39} (NCSCD) in our SOC estimates for northern latitudes^{38,39}. We generated a set of global SOC estimates, which included factorial combinations of SOC from the HWSD, extrapolated to full soil depth using both empirical models, and also used the NCSCD data set for northern latitudes. These data sets were aggregated from $\sim 1 \text{ km}^2$ (0.01° by 0.01°) to $\sim 55 \text{ km}^2$ (0.5° by 0.5°) resolution. Our global estimate of total soil organic carbon was $2,397^{+860}_{-561} \text{ PgC}$ (mean; upper limit, percentile 97.5; lower limit, percentile 2.5) (Supplementary Fig. 2). The global soil carbon stocks are comparable to a previous estimate of total soil organic carbon of $2,344 \text{ PgC}$ in the top 3 m of soil³⁴ (Supplementary Table 1). The range of estimated SOC values varied significantly between the different biome types across the world (Supplementary Table 1). According to our estimates, tropical biomes (forest, savannahs, grasslands) together account for 32% of the global soil carbon stock, and the areas of largest integrated stock are found in tropical (20%) and boreal (19%) forests.

A comprehensive assessment of the uncertainties in the SOC estimates should integrate the uncertainties from several sources: from (1) the uncertainties in the soil profiles and depth information to (2) uncertainties in the spatial extrapolation to a global extent, including (3) uncertainties in the extrapolation to full depth and (4) those emerging from the different data sources considered here (the HWSD and the NCSCD). The ability to quantify the uncertainties stemming from all these sources, and propagate them to the final SOC estimates, is limited by the available information. But here we are able to explicitly propagate the uncertainties that stem from the methods used to extrapolate SOC to full depth (from both the aforementioned empirical models; equations (2) and (3)) and from the different data sources (HWSD and NCSCD), by creating individual SOC estimates, which are combined individually with the estimates of vegetation stocks and GPP for explicit propagation of the uncertainties in τ . The quantification of the uncertainties in the total SOC stemming from depth are propagated by exploring the variability present in the soil depth from the WISE data set for the same soil types as in the HWSD. We do so by contrasting soil depth standard deviations against soil depth ($\sigma_{Df} = 0.19D_f$; $r = 0.28$, $P < 10^{-10}$, $N = 4,790$), for which we draw additionally 50 random samples of depth (with a σ_{Df} of 19%) to estimate the uncertainties in total SOC that may stem from the soil depth considered.

Overall, this approach is based on the best available information, and addresses uncertainties by evaluating the spread in the generated ensembles. We acknowledge that the integration of uncertainties stemming from the soil profiles and depth, and

from the regionalization to the global scale, could alter the uncertainties presented here. The provision of information on these sources of uncertainty, and the ability to tackle them in future estimates is essential to a more comprehensive assessment of data uncertainty.

Deriving total vegetation carbon. Our global estimates of total vegetation carbon were derived from a collection of estimates for pan-tropical regions⁴⁰ and for northern and temperate forests⁴¹ based on radar remote-sensing retrievals⁴². Above- and below-ground biomass uncertainty for the tropical regions was propagated from errors in measurements, allometric relations, sampling and predictions⁴⁰. In the Northern Hemisphere, estimates accounted for uncertainties in the BIOMASAR GSV data, wood density data and biomass compartment data⁴¹. On regional scales, the Northern Hemisphere biomass map in comparison with inventory-based data showed strong agreement (Russia: $r^2 = 0.78$ and NRMSE = 0.35; United States: $r^2 = 0.90$ and NRMSE = 0.32; Europe: $r^2 = 0.70$ and NRMSE = 0.40; NRMSE denotes the root mean squared error divided by the mean of the observations) and can thus be considered a very suitable product at 0.5° resolution. Evaluation results for the United States and Europe have shown that this data set might slightly underestimate high carbon densities due to the use of C-band radar data, but there was no systematic error detected in the intercomparison in Russia⁴¹.

One shortcoming of the above two products is the sole consideration of tree forms in their estimates. Therefore, to account for the herbaceous biomass in our estimates, we assumed a mean turnover time of one year in the live vegetation fraction per grid cell, and, given that the costs of autotrophic respiration vary significantly⁴³, we took the respiratory costs (α) to lie in the range 25%–75% (uniformly distributed):

$$C_H = \text{GPP}(1 - \alpha)f_H$$

Here C_H is the herbaceous component of carbon in vegetation; GPP is the gross primary production, based on the newest data driven estimates⁴⁴; α is the respiration cost; and f_H is the fraction of each 0.5° grid cell considered as herbaceous in the SYNMAP⁴⁵. The correlation (r^2) between the data set accounting for non-woody stocks and the original vegetation stock estimates was 0.98 ± 0.015 , and the normalized mean absolute error was 0.07.

By accounting for herbaceous cover in our global C stock estimates, we obtained differences of less than 1% in the mean global τ , with the highest differences observed for croplands (1.3%), temperate grasslands and shrublands (0.6%), and wetlands (0.6%). Therefore, accounting for herbaceous plants in our global carbon stocks did not make much of a difference to the final estimates of τ . We note that these differences may not fully reflect the dynamics in natural vegetation types, which may include below-ground perennial roots or rhizomes. However, even a threefold increase in vegetation mean residence times of carbon would result in a difference of less than 2% in total ecosystem turnover times globally. These results reflect the large contribution of woody vegetation and, mostly, of soil organic carbon to the global carbon stock estimates. Overall, terrestrial vegetation holds about $442 \pm 146 \text{ PgC}$ (mean \pm standard deviation), which is $\sim 16\%$ of the global organic carbon estimated on land (Supplementary Table 2). Excluding herbaceous vegetation, our estimate of $429 \pm 144 \text{ PgC}$ in forests encompasses the latest rounded estimate of 300 PgC derived from global inventory data⁴⁶. The most significant part of vegetation carbon is found in tropical forests and tropical savannahs and grasslands (62%), and temperate and boreal forests, and temperate grasslands and shrublands incorporate circa 25%.

Global carbon estimates. The ensemble of vegetation carbon pools was composed of 200 members, assuming normally distributed uncertainties in the satellite-derived C stocks and in GPP estimates, and a uniform distribution of autotrophic respiration costs (see previous Methods section). Each member of the ensemble of total soil carbon was individually added to each of the members of vegetation carbon ensemble. We randomly sampled 200 members of this ensemble to achieve the final data ensemble of total ecosystem carbon stocks (Fig. 1a) with uncertainties (Extended Data Fig. 1). The uncertainty bands in the latitudinal profiles in Extended Data Fig. 7 report the 5th and 95th percentiles of the data ensemble per latitudinal window.

Global mean turnover times of carbon. Turnover time is commonly defined as the ratio between the total size of a reservoir and its outflux⁴⁷. For terrestrial ecosystems, the total reservoir size is equal to the carbon stock in vegetation and soils, and the outflux comprises all carbon losses (respiration of autotrophic plants, respiration of heterotrophic organisms, losses by fire and harvest). Under the assumption that the ecosystem is neither gaining nor losing carbon (steady state), the turnover time can equivalently be calculated as the ratio between the carbon stock in vegetation (C_{veg}) and soils (C_{soil}), and the flux into this reservoir, GPP:

$$\tau = \frac{C_{veg} + C_{soil}}{\text{GPP}} \quad (4)$$

In steady state, τ is the average time that newly assimilated carbon spends in terrestrial ecosystems before it is respired, burnt or harvested. Acknowledging that our definition of τ (equation (4)) hinges on the steady-state assumption, we call τ the

apparent whole-ecosystem turnover time and interpret the quantity as an emergent diagnostic at ecosystem level (Supplementary Information Section 1). An ensemble of apparent whole-ecosystem turnover times (τ) was obtained by applying equation (1) to a random permutation of the mean annual GPP (1982–2005) and total ecosystem carbon data sets ($N = 200$). The resulting uncertainties had wide ranges in space (Supplementary Fig. 3) and between biomes (Table 1). The uncertainty bands in the latitudinal profiles in Fig. 2a report the 5th and 95th percentiles of data ensemble per latitudinal window.

Benchmarking our current results against other observation-based estimates of global carbon turnover time is hampered by the fact that previous studies have mostly focused on soils. Global turnover times of carbon in soils were reported to range from 27 yr (ref. 48) to 32 yr (ref. 49), and were generally considered to lie between 30 and 40 yr, assuming strong variations for different ecosystems⁵⁰. The spatial variations in the residence times of carbon in soils also reflect climate controls, exhibiting longer residence times in cold biomes at high latitudes (as shown in ref. 51 using ¹⁴C for forest soils). Such spatial variations are also seen in τ (Fig. 1a), which shows a latitudinal range from 255 yr in high northern latitudes (mean τ north of 75° N) to 15 yr in the equatorial tropics (Fig. 2a). In ref. 13, several estimates of global NPP (table 13), soil organic carbon (table 14) and vegetation stocks (table 12) were reported. Assuming respiration costs of around 50%, whole-ecosystem turnover time can be estimated as $\tau = C_{\text{total}}/(\text{NPP} + \text{RA})$ (RA, autotrophic respiration). By combining the different quantities of stocks and fluxes, a global τ of 21 ± 7 yr (mean \pm standard deviation) can be estimated.

For model evaluation, the separation between the soil and vegetation components of τ is useful for a more detailed diagnostic of model performance. We define τ_{soil} as the ratio between C_{soil} and net primary production (NPP) and τ_{veg} as the ratio between C_{veg} and GPP. However, the spatial representation of observation-based empirical estimates of NPP is still hampered by the difficulty of accounting for autotrophic respiration fluxes. Hence, the confidence level for NPP spatial estimates (for example, from ref. 52) still falls short to the levels of GPP estimates⁴⁴. An approach is to consider an ensemble of NPP fields supported by observation-based empirical estimates based on climate patterns. A set of state-of-the-art global NPP fields from refs 52–57 is used to build a data-driven empirical ensemble. Each of these NPP fields is combined individually with the C_{soil} ensemble members to build an ensemble of τ_{soil} for comparison to the CMIP5 model results. For the construction of τ_{veg} ensemble, we relied on the GPP and vegetation carbon stocks described above.

The data used here can be obtained from <http://www.bgc-jena.mpg.de/geodb/BGI/tau.php>.

Climate data. Climate data are based on the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim reanalysis product⁵⁸ and have been bias-corrected as described in ref. 59. We obtained daily data with a grid cell size of 0.5° during 1979–2010 from ECMWF. This reanalysis data was bias-corrected against the WATCH forcing data⁶⁰ using an overlapping period of 1979–2001 and following the standard procedure⁶¹. Ideally, this approach conserves statistical moments of the distribution, for example the mean and variance. In addition, the number of rainy days remains unchanged. The WATCH forcing data serves as the reference data set because it has already been bias-corrected against other climatic data sets⁶⁰. For the analysis, we computed the mean fields of each of the climate variables by averaging the whole data sets per grid cell between 1982 and 2005.

Correlation analysis. The association between τ and climate is assessed locally for each grid cell of the global data sets by computing local correlations using a 5.5°-by-5.5° moving window (11 by 11 grid cells). This approach enables the assessment of the local importance of climate factors (Supplementary Information Section 2) and yields a global estimate of regional covariation of τ with climate variables—temperature and precipitation. The association between τ and climate is determined by analysing the Pearson correlation, partial correlation coefficients and the non-parametric Spearman's rank correlation coefficient. To disentangle the relative importance of temperature and precipitation in determining the spatial patterns of τ , we (1) used the Lindeman–Merenda–Gold (LMG) method⁶², implemented in ref. 63, which allows to quantify the contribution of different correlated regressors (here temperature and precipitation) to a multiple linear regression model; and (2) quantified the changes in the residual sums of squares by removing each independent variable from a bivariate regression between temperature and precipitation: $\text{RI}_v = (\text{RSS}_v - \text{RSS}_{\text{tas,pr}})/\text{TSS}$, where RI_v is the relative importance of variable v , RSS_v is the residual sum of squares of the regression of τ with variable v , $\text{RSS}_{\text{tas,pr}}$ is the residual sum of squares of the regression of τ against temperature and precipitation, and TSS is the total sum of squares. Finally, the metrics (RI_{tas} and RI_{pr}) are normalized (divided by r^2) to sum to 1 (ref. 63). In addition, we performed a conditional independence test on rejecting the null hypothesis that τ is independent of precipitation or temperature given the dependence on temperature or, respectively, precipitation⁶⁴. Latitudinal correlations between τ and climate variables are based on partial correlation coefficients between τ and temperature or precipitation, controlling for

precipitation or, respectively, temperature. The analysis is conducted on a common grid size for the CMIP5 models (see next section) and the observation-derived τ and climate for a latitudinal window of $\sim 9.5^\circ$ (5 grid cells). Partial correlations are computed individually for each ensemble member of the data and the models. The uncertainty bands in the data ensemble (Fig. 2c, d) represent the 5th and 95th percentiles of the partial correlations per latitudinal window.

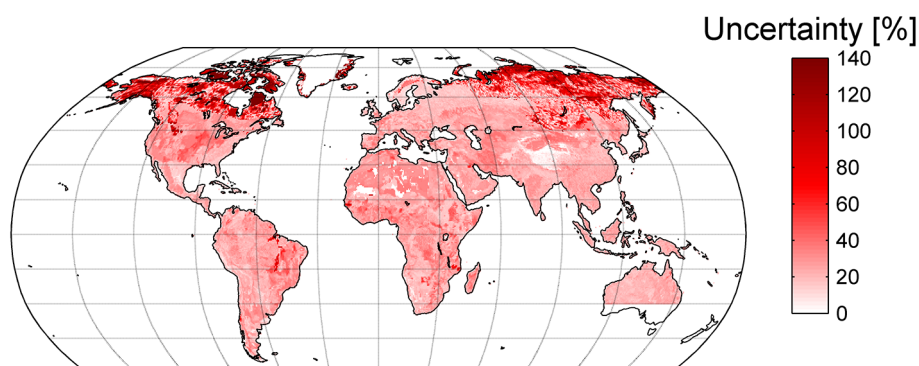
Processing Earth system model outputs from CMIP5. We analysed historical simulations outputs from ten Earth system models from CMIP5⁶⁵ (Supplementary Table 3). The historical scenario simulations (also known as the 20th-century simulations) for CMIP5 were carried out for the period from the start of the industrial revolution to near present: 1850–2005. The Earth system model here is the atmosphere–ocean coupled global climate model coupled to a carbon-cycle model, and was forced in diagnostic mode by observed changes in atmospheric composition from natural and anthropogenic sources, volcanoes, greenhouse gases and aerosols, as well as changes in solar output and land cover. The model outputs evaluated relate to climate (temperature (tas), precipitation (pr), net and shortwave downward radiation (Rn and rsds, respectively)); carbon fluxes (net ecosystem exchange (nee), gross primary production (gpp), net primary production (npp) and autotrophic and heterotrophic respiration fluxes (ra and rh, respectively) to determine ecosystem respiration ($\text{reco} = \text{ra} + \text{rh}$); and carbon pools (accounting for leaf (cLeaf), wood (cWood) and roots (cRoot) in vegetation; and accounting for soil (cSoil), litter (cLitter) and woody debris (cCwd) in soil).

The spatial fields of the variables were obtained by computing mean annual values between 1982 and 2005. The ranges stand for the common period between data availability for GPP fluxes⁴⁴ and the historical runs from CMIP5. Like for the data, the modelled values of τ are estimated from equation (1) using these simulation outputs.

Model outputs were always processed at the native spatial resolutions. To perform comparisons between models and between models and data, we constructed a common grid model ensemble. The common model ensemble was built by aggregating all model outputs to a common spatial grid, corresponding to the resolution of the NorESM1-M model ($\sim 1.89^\circ$ by 2.5° , latitude by longitude). The aggregation consisted of computing an area-weighted mean per grid cell. Because each model shows a different number of model realizations (number of model ensembles in Supplementary Table 3), we averaged all ensembles per model to avoid overweighting models with a higher number of realizations. The original grid outputs were used for within-model evaluations (partial correlations, latitudinal gradients, global and biome statistics).

33. FAO/IIASA/ISRIC/ISSCAS/JRC. *Harmonized World Soil Database v 1.2* <http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/> (2012).
34. Jobbágy, E. G. & Jackson, R. B. The vertical distribution of soil organic carbon and its relation to climate and vegetation. *Ecol. Appl.* **10**, 423–436 (2000).
35. Schumpf, M., Schulze, E. D., Kaiser, K. & Schumacher, J. How accurately can soil organic carbon stocks and stock changes be quantified by soil inventories? *Biogeochemistry* **8**, 1193–1212 (2011).
36. Webb, R. W., Rosenzweig, C. E. & Levine, E. R. *Global Soil Texture and Derived Water-Holding Capacities* (Webb et al.) http://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=548 (Oak Ridge National Laboratory Distributed Active Archive Center, 2000).
37. Zobler, L. *A World Soil File for Global Climate Modelling*. Report No. 87802 (NASA Goddard Institute for Space Studies, 1986).
38. Hugelius, G. et al. The Northern Circumpolar Soil Carbon Database: spatially distributed datasets of soil coverage and soil carbon storage in the northern permafrost regions. *Earth Syst. Sci. Data* **5**, 3–13 (2013).
39. Tarnocai, C. et al. Soil organic carbon pools in the northern circumpolar permafrost region. *Glob. Biogeochem. Cycles* **23**, GB2023 (2009).
40. Saatchi, S. S. et al. Benchmark map of forest carbon stocks in tropical regions across three continents. *Proc. Natl Acad. Sci. USA* **108**, 9899–9904 (2011).
41. Thurner, M. et al. Carbon stock and density of northern boreal and temperate forests. *Glob. Ecol. Biogeogr.* **23**, 297–310 (2014).
42. Santoro, M. et al. in *Proc. ESA Living Planet Symp.* SP-722 (CD-ROM, ESA Communication Office, 2013).
43. Amthor, J. S. The McCree-de Wit-Penning de Vries-Thornley respiration paradigms: 30 years later. *Ann. Bot. (Lond.)* **86**, 1–20 (2000).
44. Jung, M. et al. Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. *J. Geophys. Res. Biogeosci.* **116**, G00J07 (2011).
45. Jung, M., Henkel, K., Herold, M. & Churkina, G. Exploiting synergies of global land cover products for carbon cycle modeling. *Remote Sens. Environ.* **101**, 534–553 (2006).
46. Kauppi, P. E. New, low estimate for carbon stock in global forest vegetation based on inventory data. *Silva Fenn.* **37**, 451–457 (2003).
47. Rodhe, H. in *Global Biogeochemical Cycles* (eds Butcher, S. S., Charlson, R. J., Orians, G. H. & Wolfe, G. V.) 55–72 (Academic, 1992).
48. Jenkinson, D. S. in *Russell's Soil Conditions and Plant Growth* (ed. Wild, A.) 564–607 (Longman Scientific and Technical, 1988).
49. Schlesinger, W. H. in *Soils and Global Change* Vol. 25 (eds Lal, R., Kimble, J., Levine, E. & Stewart, B. A.) 9–25 (CRC/Lewis Publishers, 1995).

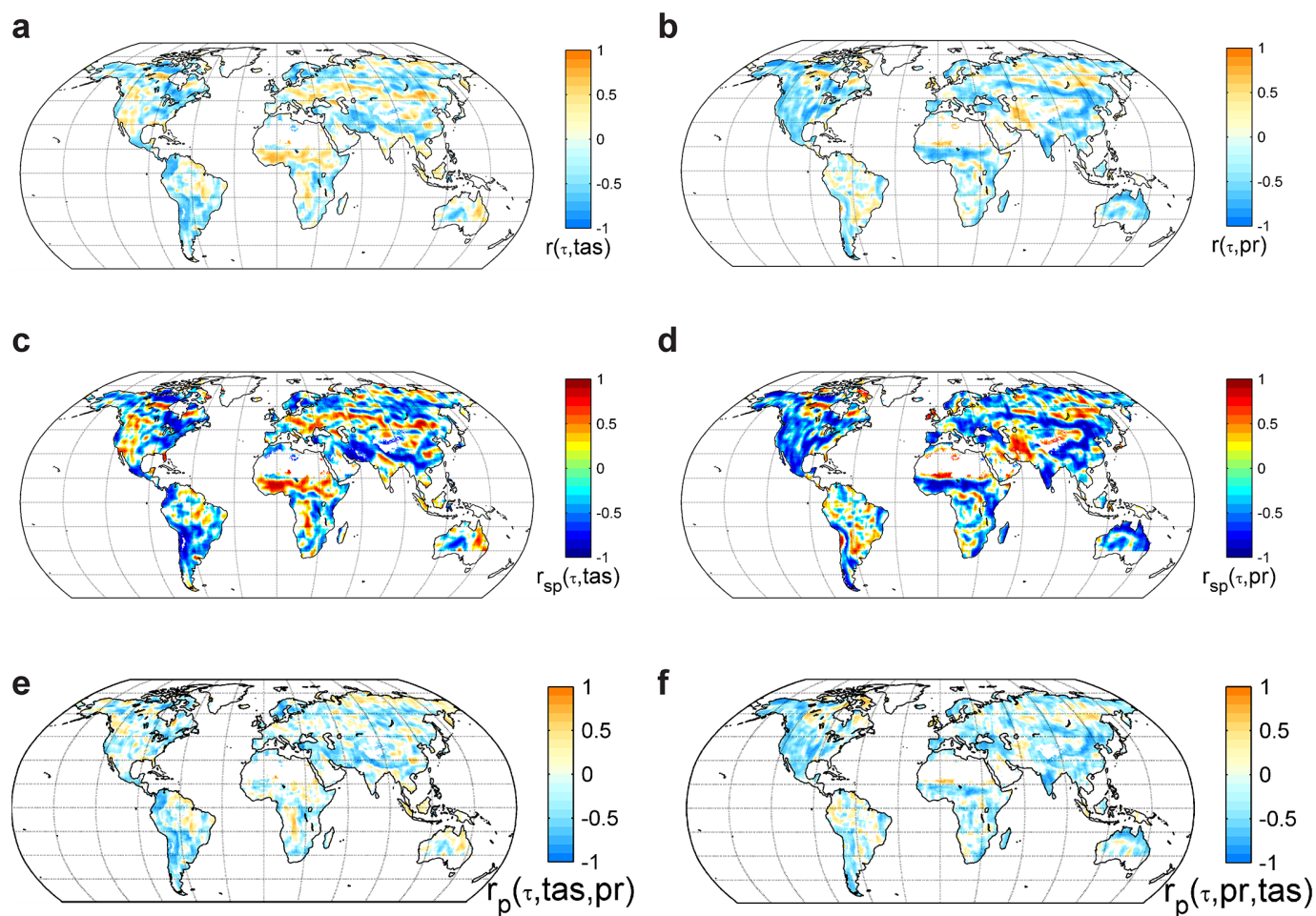
50. Oades, J. M. The retention of organic-matter in soils. *Biogeochemistry* **5**, 35–70 (1988).
51. Bird, M. I., Chivas, A. R. & Head, J. A latitudinal gradient in carbon turnover times in forest soils. *Nature* **381**, 143–146 (1996).
52. Ito, A. A historical meta-analysis of global terrestrial net primary productivity: are estimates converging? *Glob. Change Biol.* **17**, 3161–3175 (2011).
53. Zaks, D. P. M., Ramankutty, N., Barford, C. C. & Foley, J. A. From Miami to Madison: investigating the relationship between climate and terrestrial net primary production. *Glob. Biogeochem. Cycles* **21**, GB3004 (2007).
54. Del Grosso, S. *et al.* Global potential net primary production predicted from vegetation class, precipitation, and temperature. *Ecology* **89**, 2117–2126 (2008).
55. Lieth, H. in *Primary Productivity of the Biosphere* (eds Lieth, H. & Whittaker, R. H.) 237–263 (Springer, 1975).
56. Lieth, H. & Box, E. in *Publications in Climatology* (ed. Thornthwaite, W.) 37–46 (C.W. Thornthwaite Associates, 1972).
57. Schuur, E. A. G. Productivity and global climate revisited: the sensitivity of tropical forest growth to precipitation. *Ecology* **84**, 1165–1170 (2003).
58. Dee, D. P. *et al.* The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137**, 553–597 (2011).
59. Beer, C. *et al.* Harmonized European long-term climate data for assessing the effect of changing temporal variability on land-atmosphere CO₂ fluxes. *J. Clim.* **27**, 4815–4834 (2014).
60. Weedon, G. P. *et al.* Creation of the WATCH forcing data and its use to assess global and regional reference crop evaporation over land during the twentieth century. *J. Hydrometeorol.* **12**, 823–848 (2011).
61. Piani, C. *et al.* Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models. *J. Hydrol. (Amst.)* **395**, 199–215 (2010).
62. Lindeman, R. H., Merenda, P. F. & Gold, R. Z. *Introduction to Bivariate and Multivariate Analysis* (1980).
63. Grömping, U. Relative importance for linear regression in R: the package relaimpo. *J. Stat. Softw.* **17**, 1–27 (2006).
64. Zhang, K., Peters, J., Janzing, D. & Schölkopf, B. Kernel-based conditional independence test and application in causal discovery. *Computing Res. Repos.* (arXiv, 2012).
65. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An Overview of Cmp5 and the Experiment Design. *Bull. Am. Meteorol. Soc.* **93**, 485–498 (2012).
66. Prentice, I. C. *et al.* in *Climate Change 2001: The Scientific Basis* (eds Houghton, J. T. *et al.*) 183–237 (Cambridge Univ Press, 2001).
67. Beer, C. *et al.* Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate. *Science* **329**, 834–838 (2010).
68. Kottek, M., Grieser, J., Beck, C., Rudolf, B. & Rubel, F. World map of the Köppen-Geiger climate classification updated. *Meteorol. Z. (Berl.)* **15**, 259–263 (2006).
69. Turetsky, M. R. *et al.* The resilience and functional role of moss in boreal and arctic ecosystems. *New Phytol.* **196**, 49–67 (2012).
70. Page, S. E., Rieley, J. O. & Banks, C. J. Global and regional importance of the tropical peatland carbon pool. *Glob. Change Biol.* **17**, 798–818 (2011).
71. DiMiceli, C. M. *et al.* Annual Global Automated MODIS Vegetation Continuous Fields (MOD44B) at 250 m Spatial Resolution for Data Years Beginning Day 65, 2000–2010, Collection 5 Percent Tree Cover <http://glcf.umd.edu/data/vctf/> (University of Maryland, 2011).



Extended Data Figure 1 | Relative uncertainties in total ecosystem carbon.

Relative uncertainties in total ecosystem carbon stemming from the different data sources considered, reported as the ratio between the interquartile range (difference between the 75th and 25th percentiles) of the different estimates and the mean. The colour scale is binned to the 98th percentile of the spatial distribution of uncertainty (140%). A significant spatial variability was observed in the total ecosystem carbon uncertainties. The highest uncertainties locally and regarding total stocks per biome were observed in tundra ($\sim 38\%$),

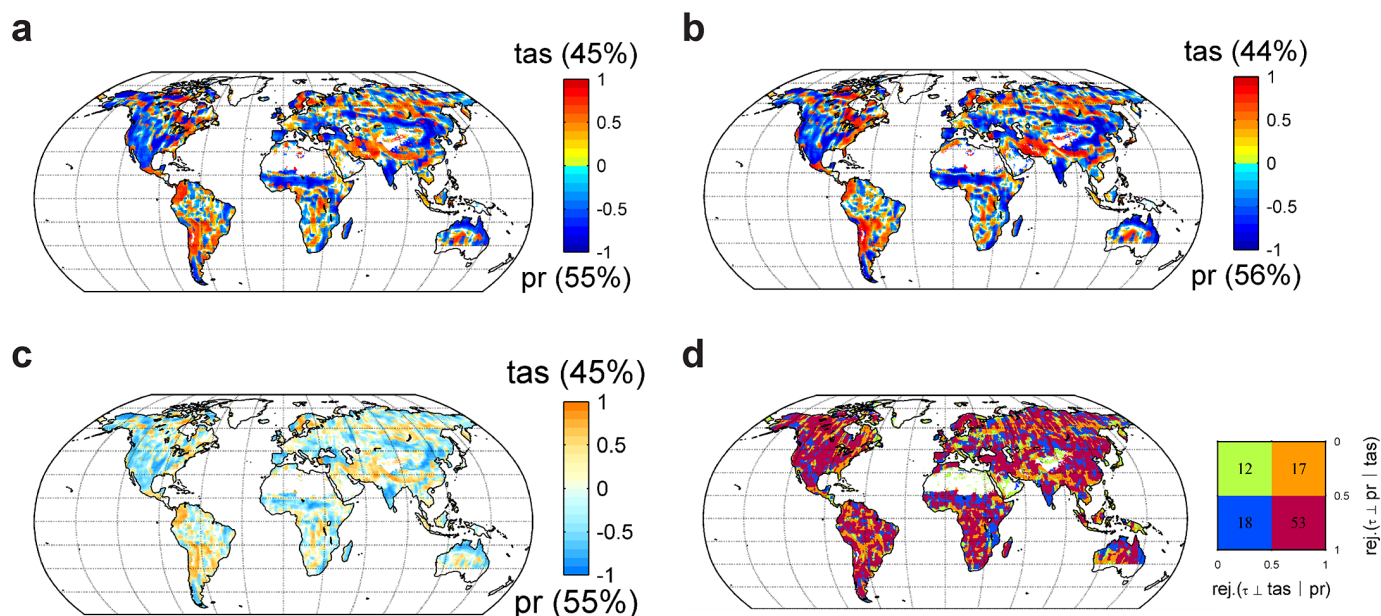
followed by tropical savannahs and grasslands ($\sim 30\%$). Deserts and croplands also showed significant relative uncertainties (both 27%). Overall, we observe a global relative uncertainty of 21% . We note unknown sources of uncertainties related to total carbon stocks, which relate mostly the representativeness of mosses in northern latitudes⁶⁹ and tropical peatlands in Southeast Asia, although we find a total soil stock of $\sim 83^{+31}_{-19}$ PgC (95% CI) in this region ($-11.5^\circ < \text{latitude} < 10^\circ$ and $90^\circ < \text{longitude} < 155^\circ$), which borders the upper envelope of the estimates in ref. 70.



Extended Data Figure 2 | Local spatial correlations between turnover times of carbon in terrestrial ecosystems and temperature, and precipitation.

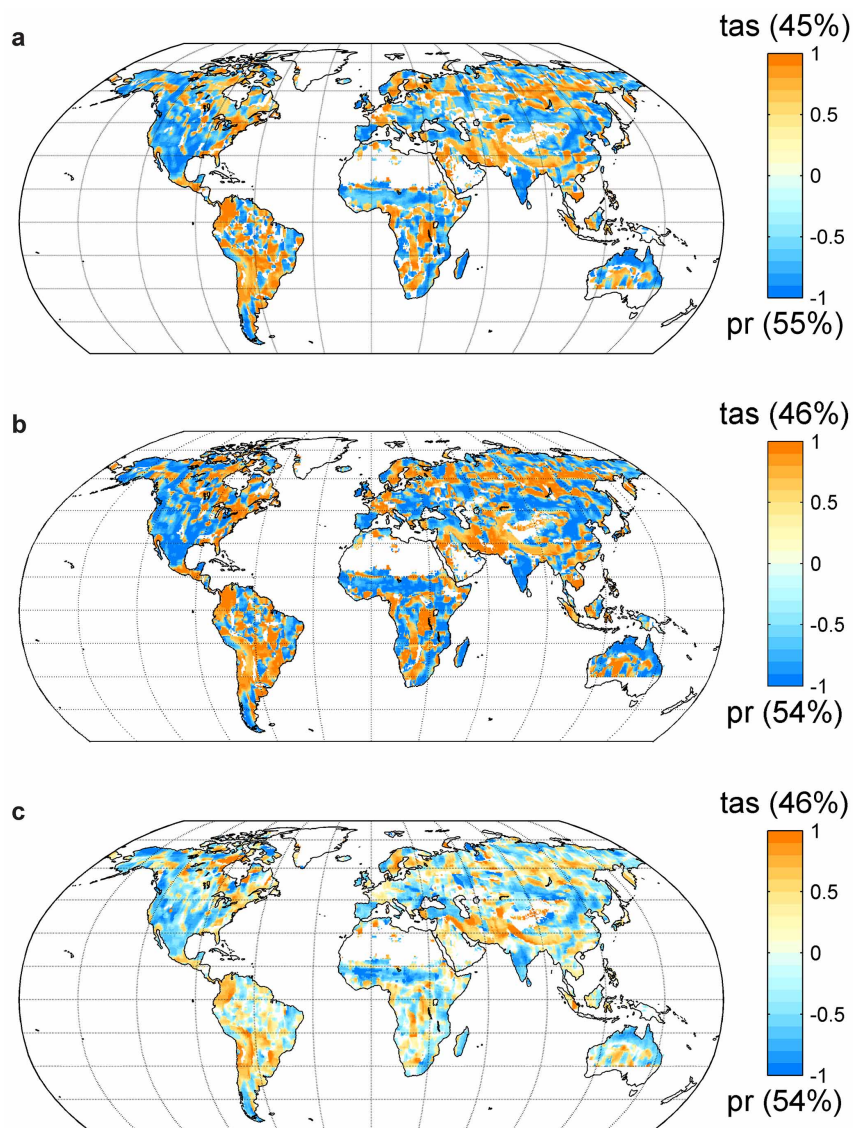
Local spatial correlations between τ and temperature (tas; **a**, **c**, **e**), and τ and precipitation (pr; **b**, **d**, **f**) using the 5.5°-by-5.5° moving-window approach. We use two alternative approaches to the Pearson correlation (**a**, **b**): the Spearman rank correlation (r_{sp}), a non-parametric measure of association that does not rely on the assumption of normal distribution of residuals (**c**, **d**);

and the partial correlation (r_p , **e**, **f**), measuring the degree of association between τ and temperature or precipitation, setting precipitation or, respectively, temperature as controlling variables (**e**, **f**). On local scales, using partial correlations may result in lost correlation owing to a strong local covariation of temperature and precipitation. Although we see this loss, the associative patterns between τ and both climate variables are generally maintained across the approaches used to calculate the correlations.



Extended Data Figure 3 | Strength of association between turnover times of carbon in terrestrial ecosystems and temperature, and precipitation, using different methods. Strength of association between τ and temperature (tas) and precipitation (pr) for Pearson correlations (**a**), Spearman correlations (**b**) and partial correlations (**c**). Each of these maps (**a–c**) shows regions where the association of τ is stronger with precipitation (blue) or temperature (red). The fraction of land grid cells with stronger significant correlations to temperature and precipitation are indicated above (for tas) and below (for pr)

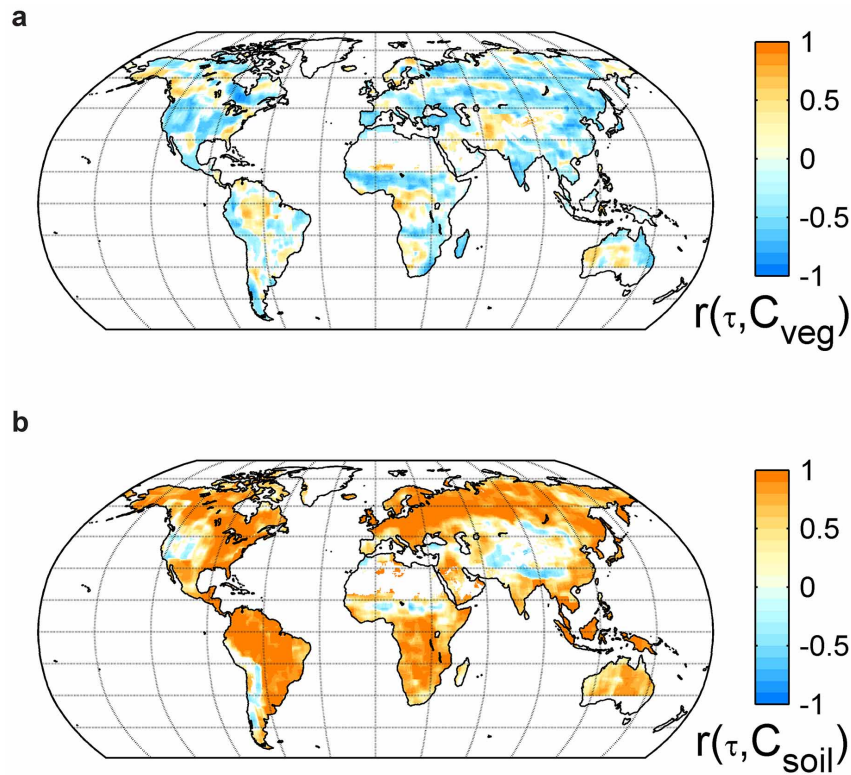
the colour bar. The colour gradients reflect the respective absolute correlation values. Despite stronger correlations with either temperature or precipitation, these cannot be said to be completely independent from the variable with lower correlation strength. **d**, Results of a conditional independence test on rejecting the null hypothesis that τ is independent from pr or tas given tas or, respectively, pr (ref. 64), showing that in 53% of the land grid cells, the dependence of τ on temperature or precipitation is not lost when controlling for precipitation or, respectively, temperature.



Extended Data Figure 4 | Maximum relative importance of temperature and precipitation in the explained variance of turnover times of carbon.

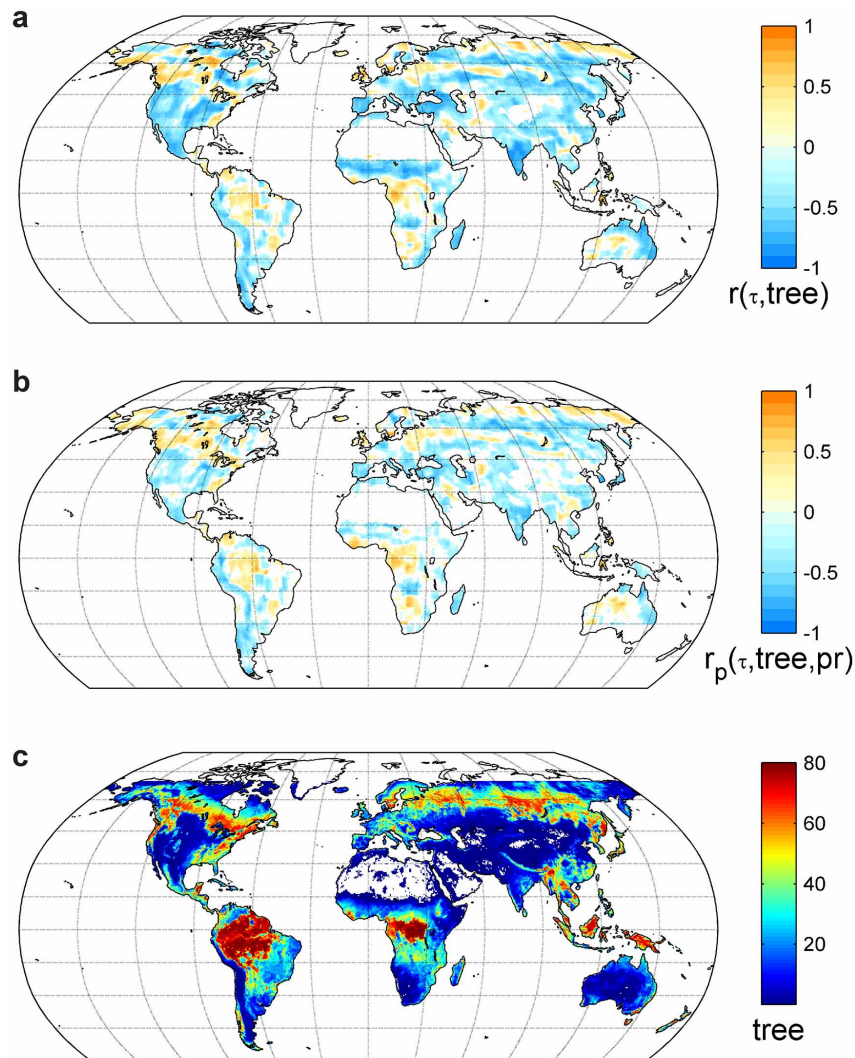
a, Maximum relative importance of temperature (tas) or precipitation (pr) in the explained variance of τ using the LMG method. **b,** Relative importance of temperature (tas) or precipitation (pr) in improving the residual sum of squares of local bivariate regressions of τ against tas and pr. **c,** Normalized slopes of the

bivariate regression between τ and precipitation and temperature, using a stepwise regression approach. Also, here the slopes correlate significantly with the strength of the association between the two variables. The fraction of land grid cells with stronger significant correlations to temperature and precipitation are indicated above (for tas) and below (for pr) the colour bar.



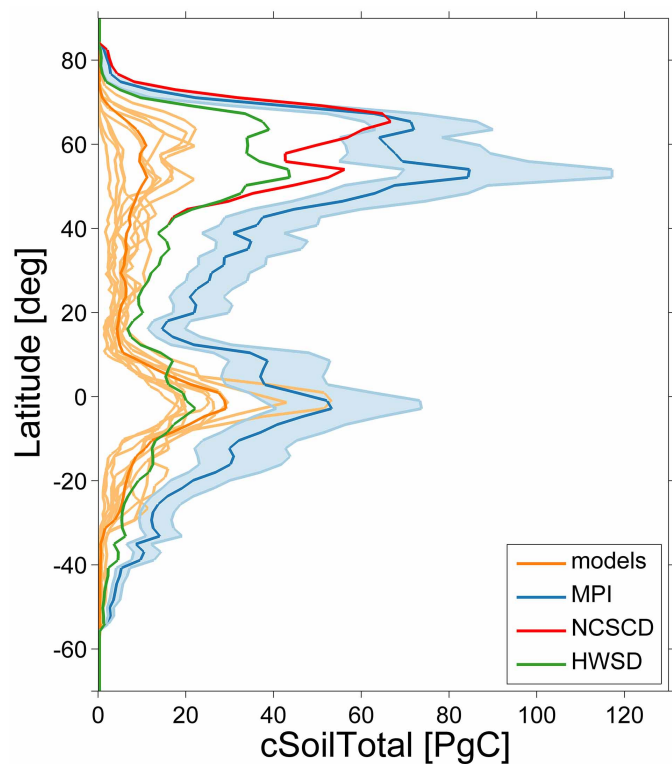
Extended Data Figure 5 | Moving-window correlation between turnover times of carbon in terrestrial ecosystems and vegetation, and soil carbon stocks. Moving-window correlation between τ and vegetation stocks (**a**); and between τ and carbon in soils (**b**). In general, τ correlates negatively with vegetation (**a**), indicating shorter turnover times with a higher proportion of carbon in the vegetation. The majority of the patterns are consistent with the

overall reduction of residence times in ecosystem carbon given allocation to vegetation pools (shorter lived by comparison with soil carbon pools). Conversely, the significance of soil carbon stocks in explaining the spatial variability of τ is pervasive (**b**). These results translate the trends in increasing τ with allocation of assimilated carbon to more persistent carbon pools.



Extended Data Figure 6 | Pearson correlations between turnover times of carbon in terrestrial ecosystems and tree cover, also controlled for the variability in precipitation. a, Pearson correlations between τ and tree cover. The prevalence of strong negative correlations suggests that the association

could be mediated by precipitation variability. **b, Controlling for precipitation still showed many of those negative correlation regions. These negative correlations are most clear in regions where tree cover is not so high or where spatial variability seems higher. c, Map of tree cover percentage from MODIS⁷¹.**



Extended Data Figure 7 | Latitudinal profiles of total soil organic carbon as simulated by CMIP5 models and from the observation-derived data ensembles. Latitudinal profiles of total soil organic carbon as simulated by CMIP5 models and from data: HWSD³³ (1 m depth), NCSCD^{38,39} (1 m depth) and this study (MPI, to full soil depth).

Extended Data Table 1 | Estimates of total ecosystem carbon for the globe and discriminated per biome

Biome type	Total stocks (PgC)			Carbon density (kgC m ⁻²)		
	Mean	P 2.5	P 97.5	Mean	P 2.5	P 97.5
Tropical forests	702	573	899	35.0	28.6	44.8
Temperate forests	292	235	382	23.4	18.8	30.6
Boreal forests	505	431	696	34.2	29.2	47.1
Tropical savannahs and grasslands	338	257	467	17.7	13.4	24.4
Temperate grasslands and shrublands	182	145	239	16.7	13.3	22.0
Deserts	250	189	346	9.4	7.1	13.0
Tundra	156	106	186	20.5	14.0	24.5
Croplands	362	279	494	23.9	18.4	32.6
Wetlands	20	15	27	21.1	16.3	28.6
Total	2807	2252	3662	22.0	17.7	28.7

Estimates of total ecosystem carbon per biome. Biomes defined according to Prentice *et al.*⁶⁶ as in Beer *et al.*⁶⁷. The ranges report the 2.5th (P 2.5) and 97.5th (P 97.5) percentiles from each ensemble member aggregated by biome.

Extended Data Table 2 | Estimates of total ecosystem carbon turnover times, stocks and fluxes of carbon for each of the CMIP5 models and correlations with data

	Global estimates								Correlation (r^2)							
Model	τ	C_{total}	C_{soil}	C_{veg}	GPP	NPP	τ_{veg}	τ_{soil}	τ	C_{total}	C_{soil}	C_{veg}	GPP	NPP	τ_{veg}	τ_{soil}
bcc-csm1-1	12.1	1507	1037	462	124	64	3.8	16.3	0.42	0.23	0.12	0.66	0.74	0.65	0.17	0.55
CanESM2	16.0	2075	1542	527	130	64	4.1	24.1	0.28	0.12	0.05	0.46	0.37	0.51	0.13	0.27
CCSM4	8.5	1101	573	522	129	45	4.1	12.6	0.17	0.22	0.03	0.65	0.75	0.77	0.14	0.42
GFDL-ESM2G	12.4	2075	1419	651	168	104	3.9	13.7	0.34	0.05	0.04	0.40	0.60	0.50	0.01	0.42
HadGEM2-ES	11.3	1567	1068	496	139	73	3.6	14.6	0.30	0.22	0.07	0.68	0.76	0.64	0.18	0.15
inmcm4	16.6	2277	1676	594	137	67	4.4	24.9	0.12	0.16	0.08	0.53	0.81	0.79	0.04	0.46
IPSL-CM5A-MR	12.3	2029	1395	629	165	82	3.8	16.9	0.26	0.28	0.15	0.42	0.68	0.38	0.13	0.52
MIROC-ESM	22.7	2923	2560	358	129	61	2.8	42.0	0.28	0.15	0.16	0.50	0.67	0.54	0.11	0.54
MPI-ESM-LR	19.7	3374	3036	336	170	88	2.0	33.9	0.46	0.001	0.003*	0.60	0.65	0.66	0.13	0.21
NorESM1-M	8.9	1153	608	538	129	46	4.2	13.1	0.18	0.31	0.13	0.61	0.71	0.72	0.13	0.47
Data (P10)	18.8	2394	1952	437	126.0	43	3.5	33.8								
Data (mean)	22.1	2797	2352	445	126.7	54	3.5	45.8								
Data (P90)	26.9	3401	2958	452	127.2	64	3.6	59.4								
Models (P10)	8.7	1131	591	346	122.9	46	2.4	12.9								
Models (mean)	14.0	1975	1466	509	140.3	69	3.7	21.3	0.38	0.45	0.26	0.79	0.85	0.82	0.18	0.59
Models (P90)	21.3	3113	2767	639	169.2	96	4.3	38.1								
Mean difference [%]	-36	-29	-38	14	11	27	4	-54								

Estimates of total ecosystem carbon turnover times (τ), carbon stocks (C_{total} , PgC), soil stocks (C_{soil} , PgC), vegetation stocks (C_{veg} , PgC), gross primary productivity (GPP, PgC yr^{-1}), net primary productivity (NPP, PgC yr^{-1}), τ_{veg} ($C_{\text{veg}}/\text{GPP}$, yr) and τ_{soil} ($C_{\text{soil}}/\text{NPP}$, yr) for each of the CMIP5 models, and correlations with data at each native model resolution (squared correlations weighted by area with removal of 2% of outliers). Global ensemble estimates (at the common ensemble spatial grid) for data and models, including lower (10th percentile, P10) and upper (90th percentile, P90) envelopes. For correlation estimates, we removed deserts (according to ref. 68) and low GPP values (below $10 \text{ gC m}^{-2} \text{ yr}^{-1}$). Low productivity values were also removed for the global estimates of τ , τ_{soil} and τ_{veg} .

* Non-significant correlations, $P > 0.05$.

High secondary aerosol contribution to particulate pollution during haze events in China

Ru-Jin Huang^{1,2*}, Yanlin Zhang^{3,4}, Carlo Bozzetti¹, Kin-Fai Ho⁵, Jun-Ji Cao², Yongming Han², Kaspar R. Daellenbach¹, Jay G. Slowik¹, Stephen M. Platt¹, Francesco Canonaco¹, Peter Zotter¹, Robert Wolf¹, Simone M. Pieber¹, Emily A. Bruns¹, Monica Crippa^{1†}, Giancarlo Ciarelli¹, Andrea Piazzalunga⁶, Margit Schwikowski^{3,4}, Gülçin Abbaszade⁷, Jürgen Schnelle-Kreis⁷, Ralf Zimmermann^{7,8}, Zhisheng An², Sönke Szidat³, Urs Baltensperger¹, Imad El Haddad^{1*} & André S. H. Prévôt¹

Rapid industrialization and urbanization in developing countries has led to an increase in air pollution, along a similar trajectory to that previously experienced by the developed nations¹. In China, particulate pollution is a serious environmental problem that is influencing air quality, regional and global climates, and human health^{2,3}. In response to the extremely severe and persistent haze pollution experienced by about 800 million people during the first quarter of 2013 (refs 4, 5), the Chinese State Council announced its aim to reduce concentrations of PM_{2.5} (particulate matter with an aerodynamic diameter less than 2.5 micrometres) by up to 25 per cent relative to 2012 levels by 2017 (ref. 6). Such efforts however require elucidation of the factors governing the abundance and composition of PM_{2.5}, which remain poorly constrained in China^{3,7,8}. Here we combine a comprehensive set of novel and state-of-the-art offline analytical approaches and statistical techniques to investigate the chemical nature and sources of particulate matter at urban locations in Beijing, Shanghai, Guangzhou and Xi'an during January 2013. We find that the severe haze pollution event was driven to a large extent by secondary aerosol formation, which contributed 30–77 per cent and 44–71 per cent (average for all four cities) of PM_{2.5} and of organic aerosol, respectively. On average, the contribution of secondary organic aerosol (SOA) and secondary inorganic aerosol (SIA) are found to be of similar importance (SOA/SIA ratios range from 0.6 to 1.4). Our results suggest that, in addition to mitigating primary particulate emissions, reducing the emissions of secondary aerosol precursors from, for example, fossil fuel combustion and biomass burning is likely to be important for controlling China's PM_{2.5} levels and for reducing the environmental, economic and health impacts resulting from particulate pollution.

In the first quarter of 2013, China experienced extremely severe and persistent haze pollution, affecting ~1.3 million km² and ~800 million people. Measurements at 74 major cities showed that the daily average concentrations of PM_{2.5} exceeded the Chinese pollution standard of 75 µg m⁻³ (about twice that of the US EPA standard of 35 µg m⁻³) for 69% of days in January, with a record-breaking daily concentration of 772 µg m⁻³ (ref. 4). This acute pollution was accompanied by extremely poor visibility and air quality, as reflected in the aerosol optical depth (AOD; Fig. 1), and a sharp increase in respiratory diseases⁵. On a longer timescale, long-range transport of pollutants from China may affect North America, the Pacific and the Arctic, making Chinese air pollution a truly global problem^{2,9}. In response to the severe haze events of 2013, the Chinese State Council quickly released the 'Atmospheric Pollution Prevention and Control Action Plan' on 10 September 2013,

which aims to reduce PM_{2.5} by up to 25% by 2017 relative to 2012 levels, and is backed by US \$277 billion in investments from the central government⁶. Achieving this highly ambitious goal requires targeted, optimized emission control strategies. However, the factors governing PM_{2.5} concentrations in China are poorly constrained^{3,7,8}, significantly hindering such efforts. Here we combine a comprehensive set of novel and state-of-the-art offline (filter-based) analytical approaches and statistical techniques to elucidate the chemical nature and predominant sources of aerosol particles during the January 2013 severe haze pollution events in China. Our results may aid the initiation of practical measures for PM_{2.5} emission reductions. Furthermore, owing to the widespread availability of ambient filters, the measurement strategies and analysis techniques developed herein are applicable to other emerging economies or developing countries, potentially facilitating their efforts to design effective mitigation strategies.

We investigated the 2013 haze pollution events with measurements at urban locations in Beijing, Shanghai, Guangzhou and Xi'an, located respectively in the northern, eastern, southern and western regions of China (see Supplementary Information section S1 for details). Average PM_{2.5} concentrations were approximately one to two orders of magnitude higher than those observed in urban areas in the US and European countries^{3,10}. As shown in Fig. 1, daily average PM_{2.5} concentrations at Xi'an (345 µg m⁻³) were more than twice those of the other sites, followed by Beijing (159 µg m⁻³), Shanghai (91 µg m⁻³) and Guangzhou (69 µg m⁻³). Chemical analyses (see Supplementary Information section S2, and Supplementary Table 1) shows that organic matter (OM) constitutes a major fraction (30–50%) of the total PM_{2.5} in all cities studied here, followed by sulphate (8–18%), nitrate (7–14%), ammonium (5–10%), elemental carbon (EC, 2–5%) and chloride (2–4%) (Fig. 1). Only about 10–15% was unidentified for Beijing, Shanghai and Guangzhou, though this increases to ~35% in Xi'an owing to elevated dust concentrations (see below), most probably consisting of crustal material such as aluminium and silicon oxides⁸.

The sources of PM_{2.5} and OM are apportioned by applying two complementary bilinear receptor models—that is, chemical mass balance (CMB)¹¹ and positive matrix factorization (PMF) using the multilinear engine ME-2 (ref. 12), see Supplementary Information section S3, Supplementary Table 2, and Supplementary Figs 1–21—and a sampling algorithm (that is, pseudo Monte Carlo calculation, see Supplementary Information section S4) to an unprecedented data set. This data set includes (1) high resolution (HR) mass spectra (by analysing nebulized water-extracts of the filter samples with a high resolution time-of-flight aerosol mass spectrometer, HR-ToF-AMS¹³), (2) organic

¹Laboratory of Atmospheric Chemistry, Paul Scherrer Institute (PSI), 5232 Villigen, Switzerland. ²State Key Laboratory of Loess and Quaternary Geology (SKLLQG), and Key Laboratory of Aerosol Chemistry and Physics, Institute of Earth Environment, Chinese Academy of Sciences, Xi'an 710075, China. ³Department of Chemistry and Biochemistry, and Oeschger Centre for Climate Change Research, University of Bern, 3012 Bern, Switzerland. ⁴Laboratory of Radiochemistry and Environmental Chemistry, Paul Scherrer Institute (PSI), 5232 Villigen, Switzerland. ⁵The Jockey Club School of Public Health and Primary Care, The Chinese University of Hong Kong, Hong Kong, China. ⁶Department of Earth and Environmental Sciences, University of Milano Bicocca, Piazza della Scienza 1, Milan 20126, Italy. ⁷Helmholtz Zentrum München, German Research Center for Environmental Health (GmbH), Joint Mass Spectrometry Centre, Cooperation Group Comprehensive Molecular Analytics and Helmholtz Virtual Institute of Complex Molecular Systems in Environmental Health — Aerosol and Health (HICE), 85764 Neuherberg, Germany. ⁸University of Rostock, Joint Mass Spectrometry Centre, Institute of Chemistry, Analytical Chemistry, 18015 Rostock, Germany. [†]Present address: European Commission, Joint Research Centre, Institute for Environment and Sustainability, Air and Climate Unit, Via Fermi, 2749, 21027 Ispra, Italy.

*These authors contributed equally to this work.

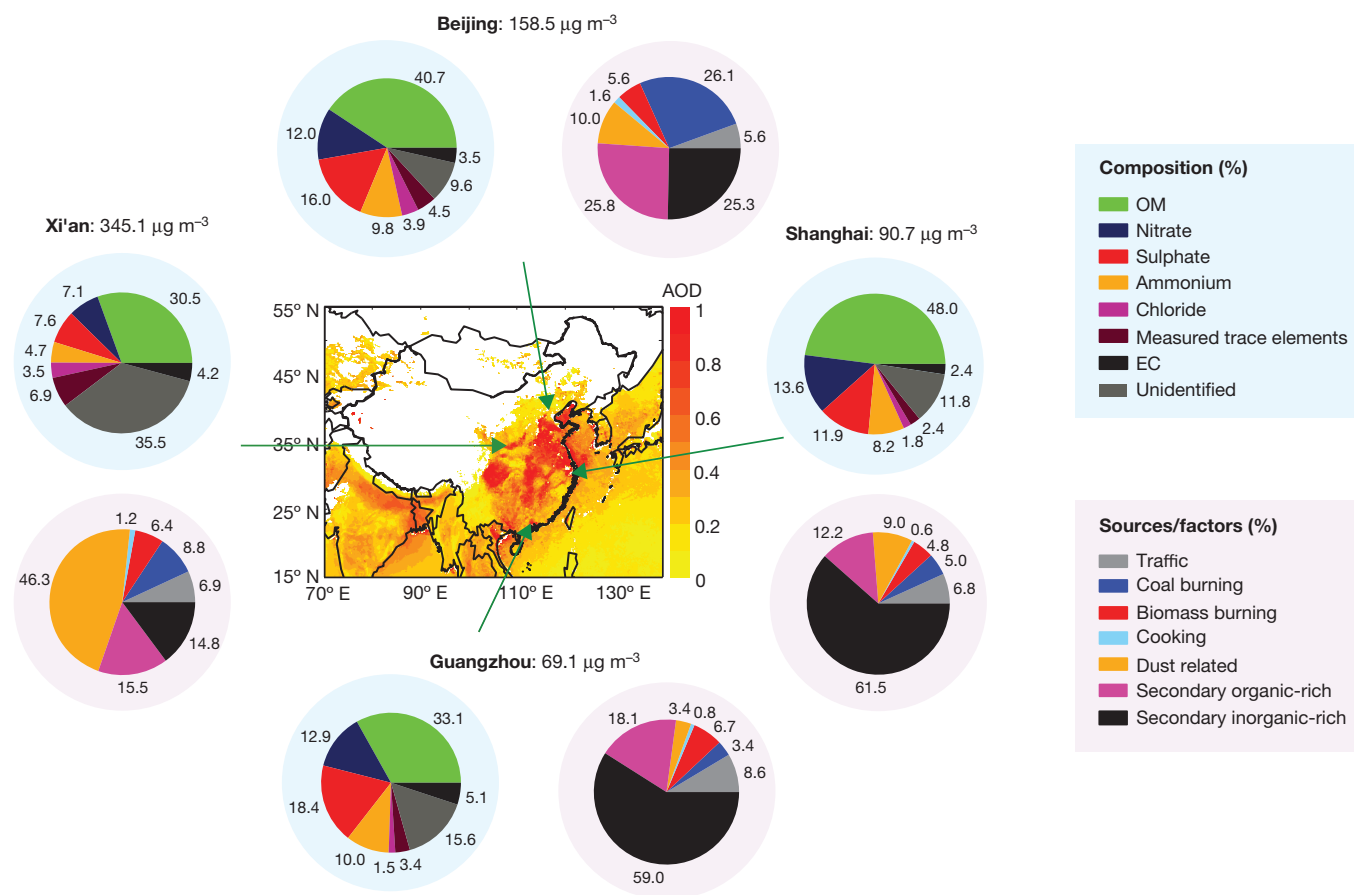


Figure 1 | Chemical composition and source apportionment of $\text{PM}_{2.5}$ collected during the high pollution events of 5–25 January 2013 at the urban sites of Beijing, Shanghai, Guangzhou and Xi'an. Centre, map showing locations of the four sites, indicated by arrows. Pie charts around the map show $\text{PM}_{2.5}$ composition and sources for each site. The measured $\text{PM}_{2.5}$ concentrations (shown next to the site name) are approximately one to two orders of magnitude higher than those observed in the urban areas of US and European countries^{3,10}. OM is the main $\text{PM}_{2.5}$ component, and secondary organic-rich and inorganic-rich aerosols are the major contributor to the $\text{PM}_{2.5}$ mass in all cities studied here except Xi'an where dust aerosol is the dominant contributor. The uncertainties (relative standard deviation, RSD) on

the source contribution estimates depend on the site considered, but are on average 24% for traffic, 39% for coal burning, 8% for biomass burning, 35% for dust related emissions, 145% for cooking, and 7% for secondary aerosols, respectively (see Supplementary Information section S4 for details). The measured trace elements include K, Na, Ca, Mg, Fe, Ti, Pb, As, Cu, Zn and Ni, while the major crustal elements, Si and Al, could not be measured due to interference from the quartz fibre substrate of the sample. The central map presents aerosol optical depth (AOD, colour-coded, see key at right), retrieved from satellite (Terra/Modis) observations over the whole month of January 2013 (<http://www.nasa.gov>), and shows the large coverage of severe particle pollution in China.

marker compounds¹⁴, (3) radiocarbon (^{14}C) content of EC and OC (organic carbon)¹⁵, (4) EC and OC, and (5) ions (see Supplementary Information section S2). The performance of the models was extensively evaluated. Model uncertainty and the sensitivity of the results to model inputs were assessed via pseudo Monte Carlo simulations (see Supplementary Information section S4, Supplementary Table 3, and Supplementary Figs 22–25). The representativeness of the measurement sites is justified in Supplementary Information section S6 and in Supplementary Fig. 28.

Critical questions for the development of pollution control strategies include both the identification of predominant sources and whether these sources are primary (that is, particles that are directly emitted to the atmosphere) or secondary (aerosol mass formed in the atmosphere from reaction products of gaseous precursors). Seven sources/factors were identified; their mean contributions are shown in Fig. 1 and factor profiles in Supplementary Fig. 17. A large fraction (51–77%) of $\text{PM}_{2.5}$ mass in Beijing, Shanghai and Guangzhou consists of secondary species, that is, SOA and SIA (the latter being sulphate, nitrate and ammonium), though this drops to ~30% in Xi'an due to higher dust levels (46% of $\text{PM}_{2.5}$ mass) in western China¹⁶. The SOA/SIA ratios were significantly higher in north China (for example, 1.4 at Xi'an and 1.3 at Beijing) than in south China (for example, 0.6 at Shanghai and 0.7 at Guangzhou). The total secondary fraction is separated into two subtypes, secondary

organic-rich and secondary inorganic-rich, reflecting differences in precursor emission patterns. The secondary organic-rich fraction correlates with the aggregate primary emissions from traffic, coal burning, biomass burning and cooking ($R^2 = 0.77$, see Supplementary Fig. 14), suggesting that this fraction is probably from oxidation products of co-emitted volatile organic compounds (VOCs), including semivolatile and intermediate volatility species¹⁷. The secondary inorganic-rich fraction correlates with the SIA species ($R^2 = 0.72\text{--}0.82$, see Supplementary Fig. 13), indicating a more regional nature⁸. Because Beijing, Shanghai, Guangzhou and Xi'an are part of larger city clusters (Beijing-Tianjin-Hebei, Yangtze River Delta, Pearl River Delta and Guanzhong city clusters, respectively), our results suggest that the transport of pollutants from the heavily populated, urbanized and industrialized surrounding areas to the city core, and probably emissions over broader geographical scales, may also contribute significantly to the formation of secondary aerosol. The differences in the contribution of secondary organic-rich/inorganic-rich fraction in each city (see Fig. 2 and more discussion below) suggest that targeted measures for SOA and SIA need to consider the effects of regional transport and local differences in emissions patterns.

Compared to the secondary fraction, the relative contribution of primary particulate emissions to $\text{PM}_{2.5}$ from individual sources is smaller: ~6–9% from traffic, 5–7% from biomass burning, 1–2% from

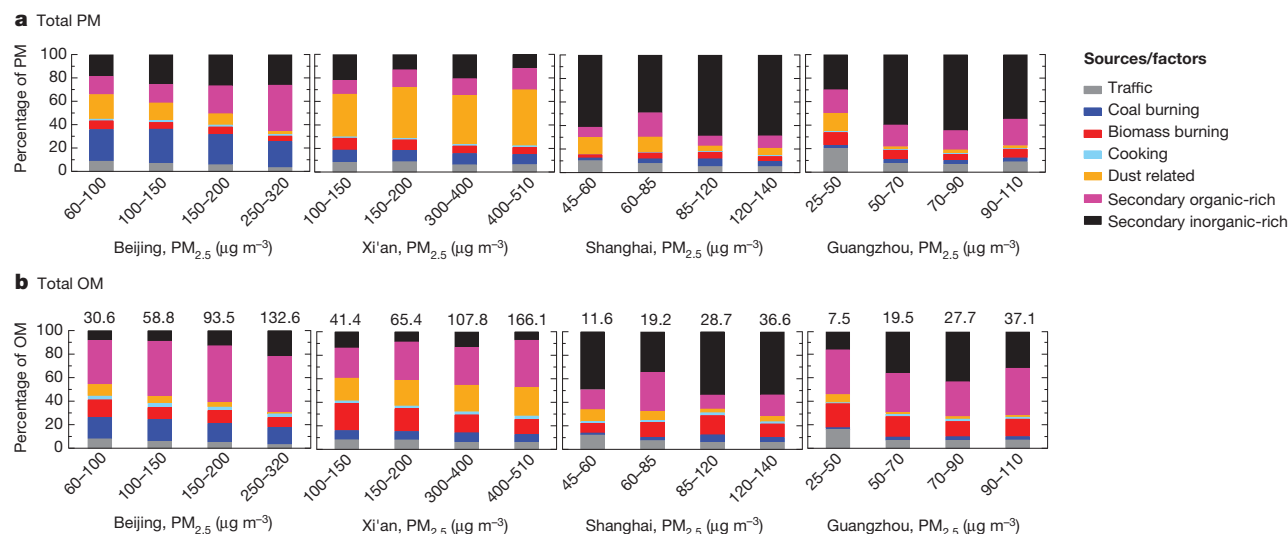


Figure 2 | Source contribution to total particulate and organic matter.

a, b, The fractional contribution to PM (**a**) and OM (**b**) as a function of the PM_{2.5} bins is shown. The uncertainties of these estimates for PM_{2.5} are given in the Fig. 1 legend. The uncertainties (RSD) for OM are on average 33% for traffic, 47% for coal burning, 8% for biomass burning, 45% for dust related emissions, and 13% for secondary aerosol (see Supplementary Information section S4 for details). The high pollution events are characterized by an increasing secondary fraction, accounting for up to 81% of PM_{2.5}

mass and 73% of OM mass. The differences in secondary organic-rich/inorganic-rich ratios of PM and OM in each city indicate the regional transport and local differences in emissions patterns. Note the high contribution from coal burning at Beijing and Xi'an due to its large usage for residential heating in wintertime, and the notably high contribution from biomass burning at Guangzhou and Xi'an. The numbers above the bars (in **b**) represent the average OM concentration in μg m⁻³.

cooking, 3–26% from coal burning and 3–10% from dust-related emission (with the exception of 46% in Xi'an) (Fig. 1). The relatively high contribution from coal burning in Beijing and Xi'an (interquartile range: 9–21% of PM_{2.5} mass (or 17–47 μg m⁻³) compared to 3–5% of PM_{2.5} mass (or 2–5 μg m⁻³) in Shanghai and Guangzhou) can be attributed to its extensive use in residential heating in northern and western China¹⁸. Our study shows that on average ~37% of sulphate is directly emitted from coal burning in Xi'an and Beijing. Coal burning emissions also dominate the levels of species associated with adverse respiratory and cardiovascular health outcomes, including polycyclic aromatic hydrocarbons and heavy metals (for example, lead and arsenic). The contribution from biomass burning is notably higher in Guangzhou and Xi'an (interquartile range: 5–9% of PM_{2.5} mass; 13–18% of OM) compared to that in Beijing and Shanghai (4–7% of PM_{2.5} mass; 8–15% of OM), consistent with previous studies¹⁹. Note that the absolute contribution from traffic in Xi'an is 2.7–4.0 times the levels at the other cities studied here, although the total vehicle fleet in Xi'an is 30–70% lower. This can be attributed to the lag in implementation of more stringent vehicle emission standards in Xi'an (see Supplementary Information section S1). The high dust levels found in Xi'an most probably originate from deserts in northwest China, consistent with back trajectory analyses (Supplementary Fig. 22), although fugitive dust from construction sites and unpaved roads could also be an important emission source, given the numerous construction activities in this region¹⁶. Measurements of crustal and anthropogenic elements by energy-dispersive X-ray fluorescence spectrometry (ED-XRF, including Fe, Ti, Ca, Zn, As, Pb, Cu, V and Ni) show that the primary sources identified above explain the levels and the variability of these elements rather well (see Supplementary Information section S5, and Supplementary Figs 26 and 27).

Figure 2a and b shows the factors driving the high pollution events by binning the fractional contribution of each factor to total PM_{2.5} and OM mass, respectively. The figures clearly show that high pollution events are characterized by an increasing secondary fraction, which accounts for up to 81% of PM_{2.5} mass and up to 73% of OM mass. On average, compared to that in the lowest PM_{2.5} bins, the secondary fraction in the highest PM_{2.5} bins increases by a factor of 1.4 for PM_{2.5}

and 1.3 for OM, demonstrating the importance of secondary aerosol formation in driving particulate pollution during high pollution events.

The significant SOA formation under wintertime conditions, though not yet widely recognized, is well supported by our low-temperature smog chamber studies on the ageing of biomass burning emissions (see Supplementary Information section S7 and Supplementary Fig. 29). We show that low temperature does not significantly reduce SOA formation rates of biomass burning emissions (the formation rate constant via OH radical (OH[•]) chemistry $k_{OH} = (2.5\text{--}6.7) \times 10^{-11} \text{ cm}^3 \text{ molecule}^{-1} \text{ s}^{-1}$ at -10°C , comparable to the oxidation rates of many SOA precursors at room temperature²⁰) and significant amounts of SOA are rapidly produced, exceeding primary organic aerosol (POA) at an OH[•] exposure (that is, OH[•] concentration \times time) of only $(1.6\text{--}5.5) \times 10^6 \text{ molecules cm}^{-3} \text{ h}$. At OH[•] concentrations typical of wintertime China ($>0.4 \times 10^6 \text{ molecules cm}^{-3}$, 24 h average, for the cities of this study), these exposures are reachable in 4–14 h, which is fast with respect to atmospheric transport especially during stagnant conditions. The OH[•] concentrations given above are simulated from the Geos-Chem model, which are consistent with wintertime observations at polluted urban locations such as Birmingham (UK), Tokyo (Japan) and New York City (see Supplementary Information section S7, Supplementary Fig. 30 for more details). Note that processes other than OH radical-initiated oxidation may contribute to or even dominate SOA formation during winter, including aqueous-phase oxidation and NO₃-radical-initiated nocturnal chemistry, as is certainly the case for SIA species sulphate and nitrate^{21,22}.

Sources of SIA are relatively well constrained: in urban areas, sulphate forms primarily through atmospheric oxidation of SO₂ emitted mainly from coal burning, while nitrate derives from NO_x emitted mainly from vehicle exhaust and power plants²³. By contrast, SOA sources are highly uncertain²⁴. We previously discussed the secondary organic-rich and secondary inorganic-rich factors in terms of likely geographic origin; here we combine the factor analysis with ¹⁴C analysis to quantify the fossil and non-fossil carbon contributions to SOA (see Supplementary Information section S4 for details). This helps to constrain the relative importance of specific sources to SOA production—for example, SOA precursors emitted by traffic/coal burning

would increase the fossil content. As shown in Fig. 3, the calculated contributions of fossil SOA to total OA mass are 1.1–2.4 times larger for high pollution events than for low pollution events, highlighting the importance of fossil SOA to particulate pollution. Fossil SOA accounts for ~25–40% of OA mass (or ~45–65% of SOA mass) in Shanghai and Beijing, consistent with the large emissions of SOA precursors from high traffic flow and/or large coal usage for domestic heating/cooking at these locations^{18,25}. The fossil SOA fraction, however, decreases to ~10–20% of OA mass in Guangzhou and Xi'an, with the non-fossil SOA fraction increasing to ~30–60% of the OA mass (or ~65–85% of the SOA mass), mostly due to the enhanced biomass burning activities as discussed above. Note that biogenic emissions may produce a small fraction of non-fossil SOA in Guangzhou owing to the relatively high temperature (5–18 °C), while this source is probably negligible in other cities of this study due to lower temperatures (that is, –12 to 2 °C at Beijing, –7 to 7 °C at Xi'an and –1 to 9 °C at Shanghai).

Our analysis suggests that emission control strategies to mitigate PM_{2.5} pollution in China should, in addition to primary particulate emissions, also address the emission reduction of secondary aerosol precursors including SO₂, NO_x, and, in particular, VOCs. The Chinese government is already making efforts to mitigate SIA precursor gases. SO₂ emissions have decreased since 2006 due to the strict nationwide implementation of flue gas desulphurization in coal-fired power plants²⁶, but NO_x continues to increase because of the lag in emission control legislations and the increase in fuel consumption by power plants and vehicles²⁶. By 2020, VOC emissions are predicted to increase by 49% relative to 2005 levels²⁷, and are not fully considered in the current Chinese air pollutant control strategy although they account for on average ~25–30% (with the exception of 16% in Xi'an) of the PM_{2.5} mass (via SOA production) and 44–71% of OM mass observed here. The importance of VOCs to controlling PM has only been recognized very recently for a few sectors, but measures are not yet well defined, being limited to the traffic and petrochemical industries⁶. However, our findings suggest that stringent controls on VOC emissions from vehicles and coal burning (introducing clean-burning stoves) could be efficient measures in regions like Beijing and Shanghai where fossil SOA formation is dominant. Control of biomass burning activities (including heating and open fires of agricultural and other waste) could be an efficient strategy in all regions, especially near Guangzhou and Xi'an

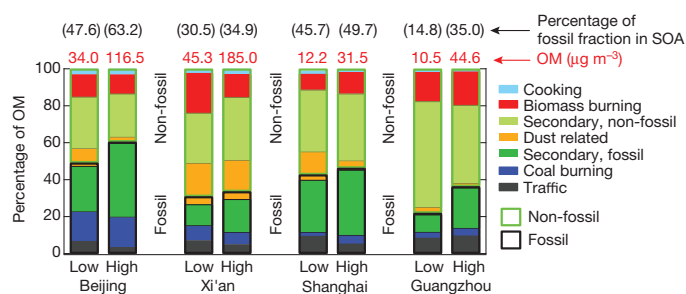


Figure 3 | Fossil and non-fossil fractional contributions of each source during low and high PM_{2.5} levels observed in different cities. For each city we measured the ¹⁴C content of carbonaceous aerosols in three samples with the lowest PM mass and three samples with the highest PM mass to represent the low and high pollution events, respectively. The numbers above the bars represent the average OM concentration and those in parentheses the percentage of the fossil fraction in SOA. The uncertainties (RSD) on the fossil and non-fossil SOA contribution to OA are on average 11% and 17%, respectively. Note that the contributions of fossil SOA to total OA mass are 1.1–2.4 times larger for high pollution events than for low pollution events, highlighting the importance of fossil SOA in particulate pollution. The fossil SOA is a dominant fraction (~25–40% of OA mass, or ~45–65% of SOA mass) in Shanghai and Beijing, consistent with the large emissions of SOA precursors from high traffic flow and large coal usage. However, non-fossil SOA is abundant in Xi'an and Guangzhou (~30–60% of OA mass, or ~65–85% of SOA mass) owing to the enhanced biomass burning activities.

where non-fossil VOC emissions are significant. Further, strategies to improve air quality within the cities need also to consider advection from extra-urban emission sources.

In general, our results suggest that reduction in SIA and SOA precursors (NO_x, SO₂, NH₃, and, particularly, the currently much less constrained VOCs) can help achieving PM_{2.5} reduction targets and diminishing the environmental, economic and health costs of particulate pollution. Such measures should be considered by policy makers given the disastrous effects of particle pollution in China. It should be investigated if similar considerations apply to other emerging economies, for example India or some African nations^{28,29}, which are also experiencing severe particle pollution. Effective air pollution control measures are imperative for cleaning the air we breathe, as highlighted by the recent World Health Organization (WHO) report attributing approximately 7 million premature deaths in 2012 to air pollution and demonstrating that particle pollution risks are far greater than previously thought³⁰.

Received 2 February; accepted 19 August 2014.

Published online 17 September 2014.

- Seinfeld, J. H. Air pollution: a half century of progress. *Am. Inst. Chem. Eng. J.* **50**, 1096–1108 (2004).
- Wang, Y., Zhang, R. Y. & Saravanan, R. Asian pollution climatically modulates mid-latitude cyclones following hierarchical modeling and observational analysis. *Nature Commun.* **5**, <http://dx.doi.org/10.1038/ncomms4098> (2014).
- Cao, J. J. Pollution status and control strategies of PM_{2.5} in China. *J. Earth Environ.* **3**, 1030–1036 (2012).
- China National Environmental Monitoring Centre. *Air Quality Report in 74 Chinese Cities in March and the First Quarter 2013* (http://www.cnemc.cn/publish/106/news/news_34605.html (in Chinese), accessed on 11 June 2013).
- Chen, R. J., Zhao, Z. H. & Kan, H. D. Heavy smog and hospital visits in Beijing, China. *Am. J. Respir. Crit. Care Med.* **188**, 1170–1171 (2013).
- Chinese State Council. *Atmospheric Pollution Prevention and Control Action Plan* (http://www.gov.cn/jwzqk/2013-09/12/content_2486773.htm (in Chinese), accessed on 12 September 2013).
- Zhang, Q., He, K. B. & Huo, H. Cleaning China's air. *Nature* **484**, 161–162 (2012).
- Yang, F. et al. Characteristics of PM_{2.5} speciation in representative megacities and across China. *Atmos. Chem. Phys.* **11**, 5207–5219 (2011).
- Wuebbles, D. J., Lei, H. & Lin, J. T. Intercontinental transport of aerosols and photochemical oxidants from Asia and its consequences. *Environ. Pollut.* **150**, 65–84 (2007).
- Jimenez, J. L. et al. Evolution of organic aerosols in the atmosphere. *Science* **326**, 1525–1529 (2009).
- Watson, J. G. et al. *CMB8 Applications and Validation Protocol for PM_{2.5} and VOCs* (US Environmental Protection Agency and Desert Research Institute, Reno, Nevada, 1998).
- Canonaco, F., Crippa, M., Slowik, J. G., Baltensperger, U. & Prévôt, A. S. H. SoFi, an IGOR-based interface for the efficient use of the generalized multilinear engine (ME-2) for source apportionment: ME-2 application to aerosol mass spectrometer data. *Atmos. Meas. Tech.* **6**, 3649–3661 (2013).
- DeCarlo, P. F. et al. Field-deployable, high-resolution, time-of-flight aerosol mass spectrometer. *Anal. Chem.* **78**, 8281–8289 (2006).
- Orasche, J., Schnelle-Kreis, J., Abbaszade, G. & Zimmermann, R. Technical note: in-situ derivatization thermal desorption GC-TOFMS for direct analysis of particle-bound non-polar and polar organic species. *Atmos. Chem. Phys.* **11**, 8977–8993 (2011).
- Zhang, Y. L. et al. On the isolation of OC and EC and the optimal strategy of radiocarbon-based source apportionment of carbonaceous aerosols. *Atmos. Chem. Phys.* **12**, 10841–10856 (2012).
- Cao, J. J. et al. On the potential high acid deposition in northeastern China. *J. Geophys. Res.* **118**, 4834–4846 (2013).
- Robinson, A. L. et al. Rethinking organic aerosols: semivolatile emissions and photochemical aging. *Science* **315**, 1259–1262 (2007).
- Zheng, M. et al. Seasonal trends in PM_{2.5} source contributions in Beijing, China. *Atmos. Environ.* **39**, 3967–3976 (2005).
- Wang, G. H. et al. High loadings and source strengths of organic aerosols in China. *Geophys. Res. Lett.* **33**, L22801 (2006).
- Atkinson, R. & Arey, J. Atmospheric degradation of volatile organic compounds. *Chem. Rev.* **103**, 4605–4638 (2003).
- Wang, X. F. et al. The secondary formation of inorganic aerosols in the droplet mode through heterogeneous aqueous reactions under haze conditions. *Atmos. Environ.* **63**, 68–76 (2012).
- Ervens, B., Turpin, B. J. & Weber, R. J. Secondary organic aerosol formation in cloud droplets and aqueous particles (aqSOA): a review of laboratory, field and model studies. *Atmos. Chem. Phys.* **11**, 11069–11102 (2011).
- Seinfeld, J. H. & Pandis, S. N. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change* 2nd edn (Wiley, 2006).
- Hallquist, M. et al. The formation, properties and impact of secondary organic aerosol: current and emerging issues. *Atmos. Chem. Phys.* **9**, 5155–5236 (2009).
- Wang, X. et al. Characterization of organic aerosol produced during pulverized coal combustion in a drop tube furnace. *Atmos. Chem. Phys.* **13**, 10919–10932 (2013).

26. Wang, Y., Zhang, Q. Q., He, K., Zhang, Q. & Chai, L. Sulfate-nitrate-ammonium aerosols over China: response to 2000–2015 emission changes of sulfur dioxide, nitrogen oxides, and ammonia. *Atmos. Chem. Phys.* **13**, 2635–2652 (2013).
27. Xing, J. *et al.* Projections of air pollutant emissions and its impacts on regional air quality in China in 2020. *Atmos. Chem. Phys.* **11**, 3119–3136 (2011).
28. Tiwari, S. *et al.* Diurnal and seasonal variations of black carbon and PM_{2.5} over New Delhi, India: Influence of meteorology. *Atmos. Res.* **125–126**, 50–62 (2013).
29. The United Nations Environment Program (UNEP). *Africa Environment Outlook 3: Our Environment, Our Health* (2013); available at <http://www.unep.org/pdf/aeo3.pdf>.
30. The World Health Organization (WHO). *7 Million Premature Deaths Annually Linked to Air Pollution* (published online 25 March 2014); available at <http://www.who.int/mediacentre/news/releases/2014/air-pollution/en/>.

Supplementary Information is available in the online version of the paper.

Acknowledgements The research leading to these results received funding from the European Community's Seventh Framework Programme (FP7/2007–2013) under grant agreement no. 290605, the Swiss National Science Foundation (SAPMAV,

no.200021_13016, WOOSHI, no. 200021L_140590, and Ambizione, PZ00P2_131673), the Swiss Competence Centers Environment and Sustainability as well as Energy and Mobility under project OPTIWARES, the National Science Foundation of China (no. 40925009), the “Strategic Priority Research Program” of the Chinese Academy of Sciences (XDA05100402), and the Helmholtz Virtual Institute of Complex Molecular Systems in Environmental Health – Aerosol and Health (HICE). The help of G. Salazar (University of Bern) during ¹⁴C analysis is acknowledged.

Author Contributions R.-J.H., I.E.H. and C.B. wrote the paper. R.-J.H., J.-J.C. and A.S.H.P. designed the study. R.-J.H., I.E.H., C.B. and K.R.D. performed the offline AMS analysis. Y.Z., P.Z. and S. S. performed the ¹⁴C analysis. M.S. performed the IC analysis. G.A. and J.S.-K. performed the TD-GC-MS analysis. R.-J.H., I.E.H., C.B. and A.S.H.P. analysed the data. All authors reviewed and commented on the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.-J.C. (jjcao@ieecas.cn) or A.S.H.P. (andre.prevot@psi.ch).

Pleistocene cave art from Sulawesi, Indonesia

M. Aubert^{1,2*}, A. Brumm^{1†*}, M. Ramli³, T. Sutikna^{1,4}, E. W. Saptomo⁴, B. Hakim⁵, M. J. Morwood[‡], G. D. van den Bergh¹, L. Kinsley⁶ & A. Dosseto^{7,8}

Archaeologists have long been puzzled by the appearance in Europe ~40–35 thousand years (kyr) ago of a rich corpus of sophisticated artworks, including parietal art (that is, paintings, drawings and engravings on immobile rock surfaces)^{1,2} and portable art (for example, carved figurines)^{3,4}, and the absence or scarcity of equivalent, well-dated evidence elsewhere, especially along early human migration routes in South Asia and the Far East, including Wallacea and Australia^{5–8}, where modern humans (*Homo sapiens*) were established by 50 kyr ago^{9,10}. Here, using uranium-series dating of coralloid speleothems directly associated with 12 human hand stencils and two figurative animal depictions from seven cave sites in the Maros karsts of Sulawesi, we show that rock art traditions on this Indonesian island are at least compatible in age with the oldest European art¹¹. The earliest dated image from Maros, with a minimum age of 39.9 kyr, is now the oldest known hand stencil in the world. In addition, a painting of a babirusa ('pig-deer') made at least 35.4 kyr ago is among the earliest dated figurative depictions worldwide, if not the earliest one. Among the implications, it can now be demonstrated that humans were producing rock art by ~40 kyr ago at opposite ends of the Pleistocene Eurasian world.

Sulawesi is the world's eleventh largest island and the biggest and probably oldest in Wallacea, the zone of oceanic islands between continental Asia and Australia. The Eocene to middle Miocene limestones of the Maros and Pangkep regions lie between 4° 7' S and 5° 1' S and cover an area of ~450 km² parallel to the west coast of the island's southwestern peninsula¹² (Fig. 1). Rivers draining the volcanic highlands to the east cut down into the basal limestone, forming clusters of plateau-like karst towers that rise abruptly from the surrounding alluvial plains¹². Extensive networks of footcaves were formed around the tower bases and now harbour abundant evidence of prehistoric human occupation¹³. Cemented breccia banks containing archaeological material occur on the rear walls of many caves and rockshelters^{14,15}, and at least 90 rock art sites are recorded. While multiple cave and shelter sites have been excavated since the 1930s (ref. 16), only two with Pleistocene sequences—Leang Burung 2 (ref. 13) and Leang Sakapao 1 (ref. 17)—have so far been reported (Fig. 1). The oldest, Leang Burung 2, a cliff-foot shelter with a minimum age for the excavated deposits of 31,260 ± 320 radiocarbon years BP (35,248 ± 420 calendar years BP)¹³, previously provided the earliest dated evidence for humans on Sulawesi. The Pleistocene deposits from both sites yielded evidence of pigment use in the form of faceted haematite nodules¹³ and ochre-smear stone tools¹⁷.

The Maros–Pangkep rock art was first recorded in the 1950s (ref. 15) and has been extensively studied by Indonesian researchers, although few detailed reports have been published. On the basis of superimposition, two broad periods of prehistoric art production are defined¹⁸. The earliest of these is characterized by human hand stencils (made by spraying wet pigment around hands pressed against rock surfaces) and, less commonly, large naturalistic paintings of endemic Sulawesi land

mammals, including the dwarfed bovid anoa (*Anoa* sp.), Celebes warty pig (*Sus celebensis*) and the 'pig-deer' babirusa (*Babirusa* sp.). These wild animal species are most commonly depicted in profile as irregularly infilled outlines¹⁸.

The later rock art phase in the Maros–Pangkep karsts lacks images of this nature. It is instead typified by small depictions of zoomorphs (including dogs and other domesticated species), anthropomorphs and a wide range of geometric signs, most commonly drawn onto rock surfaces using black pigment (possibly charcoal)¹⁸. This art can plausibly be attributed to early Austronesian immigrants on the basis of stylistic elements¹⁹, and is thus at most a few thousand years old²⁰.

The red and mulberry-coloured motifs of the earlier phase typically occur on high roofs, elevated parts of rock walls or other difficult-to-access areas in caves and shelters¹⁸. They are located both close to site entrances

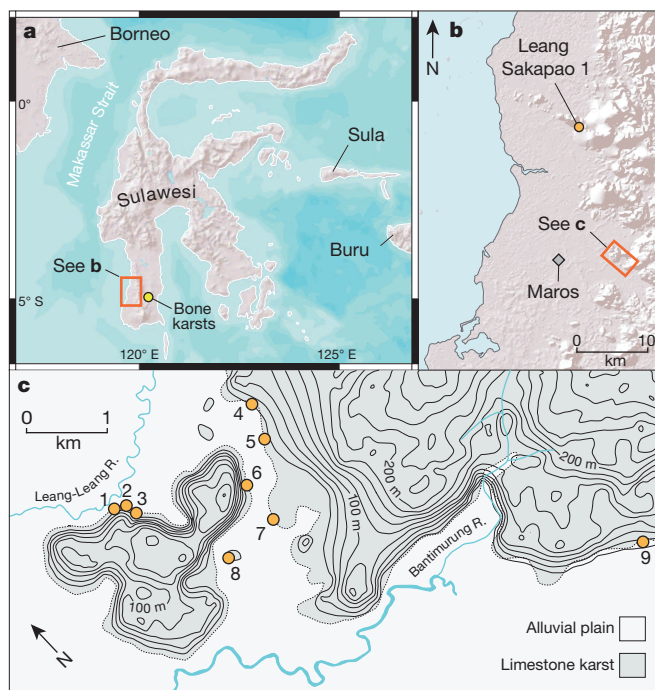


Figure 1 | Location of the study area. **a**, Sulawesi is situated east of Borneo in the Wallacean archipelago. **b**, The location of the Maros–Pangkep karsts (the area of high relief) near the town of Maros on Sulawesi's southwestern peninsula. The separate karst region of Bone is further east. **c**, The locations of the archaeological sites included in this study: 1, Leang Barugayya 2; 2, Leang Barugayya 1; 3, Gua Jing; 4, Leang Bulu Bettue; 5, Leang Sampeang; 6, Leang Timpuseng; 7, Leang Burung 2; 8, Leang Lompoa; and 9, Leang Jarie. Gua Jing and Leang Barugayya 1 and 2 are separate cave sites interconnected by a system of phreatic passages. Map data: copyright ESRI (2008).

¹Centre for Archaeological Science, University of Wollongong, Wollongong, New South Wales 2522, Australia. ²Place, Evolution and Rock Art Heritage Unit (PERAHU), Griffith University, Gold Coast, Queensland 4222, Australia. ³Balai Pelestarian Peninggalan Purbakala, Makassar 90111, Indonesia. ⁴National Centre for Archaeology (ARKENAS), Jakarta 12001, Indonesia. ⁵Balai Arkeologi Makassar, Makassar 90242, Indonesia. ⁶Research School of Earth Sciences, The Australian National University, Canberra, Australian Capital Territory 0200, Australia. ⁷Wollongong Isotope Geochronology Laboratory, University of Wollongong, Wollongong, New South Wales 2522, Australia. ⁸GeoQuEST Research Centre, University of Wollongong, Wollongong, New South Wales 2522, Australia. †Present address: Environmental Futures Research Institute, Griffith University, Brisbane, Queensland 4111, Australia.

*Deceased.

*These authors contributed equally to this work.

and within deep, dark chambers and passages. In most cases the art is poorly preserved, surviving only as weathered patches of pigment on exfoliated rock surfaces. At some sites, better-preserved art is partly or almost completely obscured by dense clusters of small coralloid speleothems ('cave popcorn') up to ~10 mm thick, which form when thin films of water precipitate on rock surfaces²¹. At one Maros cave site, Leang Bulu Bettue (Fig. 1), we observed Austronesian style drawings on a 'fresh' limestone ceiling formed by shedding of an earlier surface containing faded hand stencils (Extended Data Fig. 1), suggesting that even in recent prehistoric times this art was in an advanced state of deterioration. Despite this, local custodians report that the loss of the art has accelerated in recent decades.

To determine the age of the earliest rock art in the Maros karsts we undertook an extensive program of uranium-series dating of coralloid speleothems directly associated with the motifs. The sampled materials all comprise static coralloids that formed directly on top of clearly discernible motifs, offering the possibility to obtain minimum ages for the underlying rock art. In some cases, hand stencils and paintings were made over coralloids that then continued to grow, providing an opportunity to obtain both minimum and maximum ages for the art.

We collected a total of 19 coralloid samples associated with 14 individual motifs (12 hand stencils and 2 figurative animal depictions) (Figs 2 and 3 and Extended Data Figs 2–9) at seven cave sites in the Maros karsts (Fig. 1). Six of these sites are located within a ~1-km radius in the

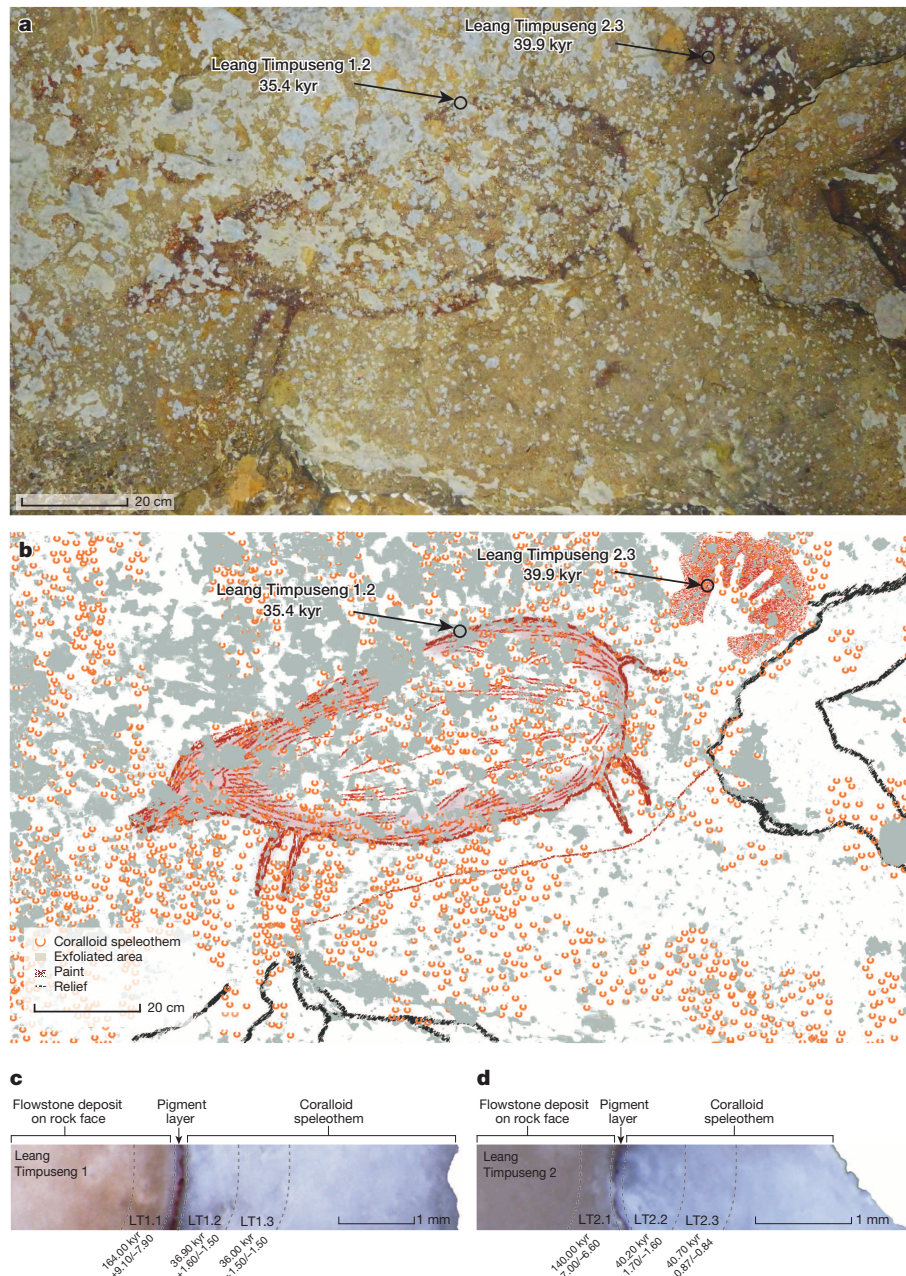


Figure 2 | Dated rock art from Leang Timpuseng. **a, b**, Photograph (**a**) and tracing (**b**) showing the locations of the dated coralloid speleothems and associated paintings: a hand stencil and a large naturalistic depiction of an animal shown in profile. Although the animal figure is badly deteriorated and obscured by coralloids, we interpret it as a female babirusa. A painted red line below the babirusa (not clearly visible in **a**, but illustrated in **b**) seems to

represent the ground surface on which the animal is standing or walking. The rock art panel is located on the ceiling about 8 m from the cave entrance and 4 m above the current cave floor. **c, d**, Profiles of the coralloid speleothems showing the microexcavated subsamples bracketing the age of the paintings. We interpret the similar ages for the overlying aliquots as a result of fast-growing speleothems. Tracing credit: Leslie Refine 'Graph & Co' (France).

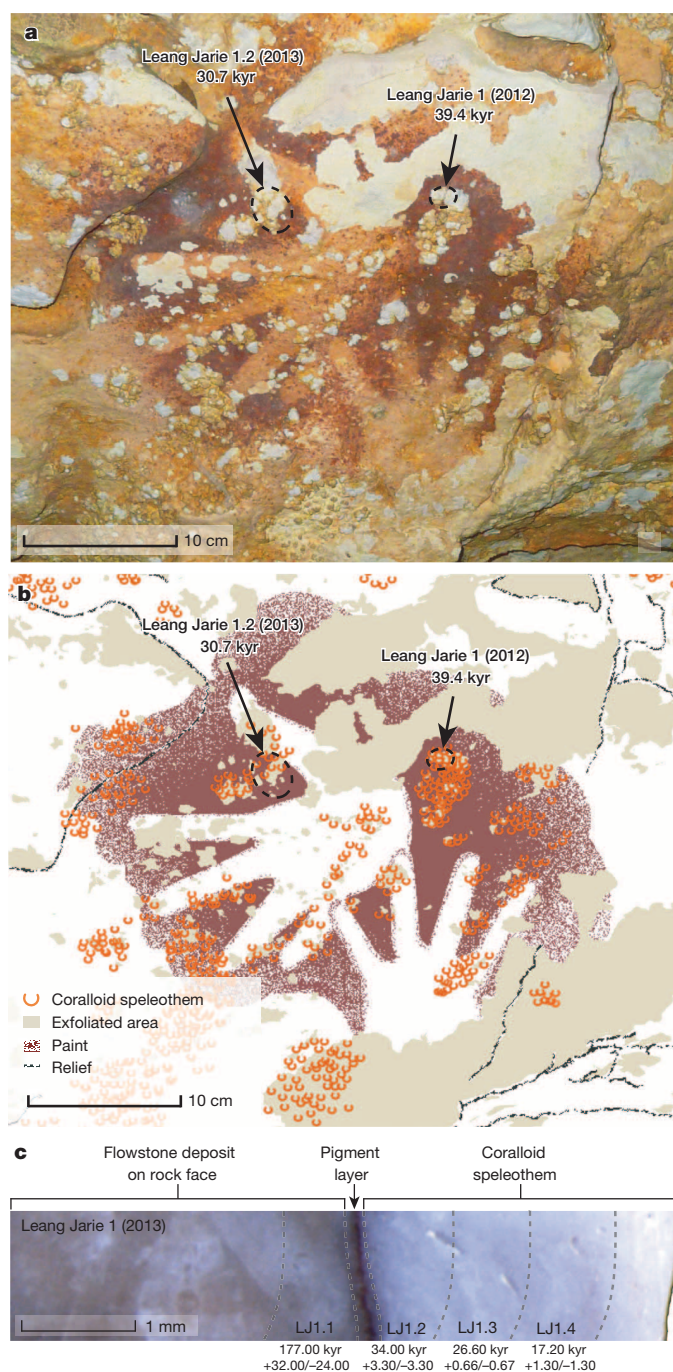


Figure 3 | Dated rock art from Leang Jarie. **a**, **b**, Photograph (**a**) and tracing (**b**) showing the locations of the dated coralroid speleothems and associated hand stencils. The hand stencils are part of a 4-m-long art panel located in a dark recess along the eastern wall of the cave, about 5 m from the entrance and 1.5 m above the floor. **c**, Profile of the coralroid speleothem (Leang Jarie 1 (2013)) showing the microexcavated subsamples bracketing the age of the paintings. The Leang Jarie 1 (2012) sample is from above the pigment layer and so only provides a minimum age for the underlying hand stencils. Tracing credit: Leslie Refine 'Graph & Co' (France).

Bantimurung region, close to Leang Burung 2. Four of the Bantimurung sites (Gua Jing, Leang Barugayya 1 and 2, and Leang Timpuseng) are situated in a large limestone outlier roughly 2 km in diameter and 180 m high¹². Leang Sampeang is located in an elevated niche on tall limestone cliffs ~500 m east of the outlier, whereas Leang Lompoa occurs at the base of an adjacent karst inselberg. The seventh cave site, Leang Jarie, is in the Simbang district southeast of Bantimurung (Fig. 1).

To provide an internal check of the microstratigraphic order of ages we took a minimum of three (and up to six) aliquots from every sample (except for Samples Leang Jarie 1 and 2 (2012)), one from under the pigment layer and two or more from above it, giving a total of 55 uranium-series age determinations (Supplementary Information). In addition, at Leang Jarie (Fig. 3), Leang Barugayya 2 (Extended Data Fig. 6) and Leang Sampeang (Extended Data Fig. 9) we dated two coralroids that had formed over the same motif. At Leang Lompoa (Extended Data Fig. 3) and Leang Jarie (Extended Data Fig. 2) we also dated two samples taken from different parts of the same coralroid. Dating results for these five sets of paired samples are internally consistent (Supplementary Information), demonstrating the robustness of the ages for the associated motifs.

Minimum ages for the Maros rock art motifs ($n = 14$) span the time range between 39.9 and 17.4 kyr ago, with the majority dating to more than 25 kyr ago (Table 1 and Supplementary Information). The oldest dated motif is a hand stencil from Leang Timpuseng, which has a minimum age of 39.9 kyr (Fig. 2) and is now the earliest evidence for humans on Sulawesi, as well as the oldest known example of this widespread art form. This motif is located on a 4-m-high ceiling next to a large irregularly infilled painting of a female babirusa, which has a minimum age of 35.4 kyr (Fig. 2). At nearby Leang Barugayya 2, a large painting of an indeterminate animal (probably a pig) has a minimum age of 35.7 kyr (Extended Data Fig. 6). The next oldest motif in our assemblage is another hand stencil at Leang Jarie, which dates to at least 39.4 kyr ago (Fig. 3).

With the Leang Timpuseng hand stencil, and for many other motifs in our sample, subsamples taken from below the pigment layer were more than 100 kyr in age (Supplementary Information). These early dates represent calcium carbonate deposits (flowstone layers) present on the rock face before the art was produced. At Gua Jing we dated two distinct hand stencils, one of which yielded minimum and maximum ages of 22.9 and 27.2 kyr, respectively (Extended Data Fig. 8). Thus, given that the Leang Timpuseng hand stencil has a minimum age of 39.9 kyr, we can infer the existence in the Maros karsts of an artistic culture with a duration of at least ~13 kyr.

The discovery of rock art dating back at least 40 kyr ago on Sulawesi has implications for our understanding of the time-depth of early symbolic traditions in the region, about which little is currently known. For instance, rock art complexes that are focused on hand stencils and large animal paintings occur in the Bone karsts ~35 km east of Maros (Fig. 1), as well as west of Sulawesi in Kalimantan (Borneo)^{22,23} and further afield in mainland Southeast Asia²⁴. The northern Australian rock art provinces of Arnhem Land²⁵ and the Kimberley²⁶ also display early art phases (based on order of superimposition) characterized by hand stencils and large irregularly infilled paintings of animals, including apparent images of extinct megafauna^{25,26}, that are markedly similar in style to the Maros art. Given that the deepest excavated deposits in northern Australia (dated to ~50–40 kyr ago) contain use-worn haematite crayons and other evidence of ochre processing and use^{9,10,27}, it is possible that an extensive archive of rock art may yet survive from the initial modern human colonization of Australia and Southeast Asia.

There are also implications for current thinking about the origins of Palaeolithic rock art, which is invariably dominated by European data and for which there are two widely debated models^{11,28}. The first of these is that rock art originated in Europe and developed gradually over thousands of years, beginning with abstract, non-figurative imagery (for example geometric patterns) and culminating in sophisticated naturalistic representations of animals, such as those in Altamira and Lascaux dated to ~20 kyr ago^{11,28,29} and other late Upper Palaeolithic cave sites in western Europe. This long-standing notion is given new impetus by recent uranium-series dating of rock art motifs from 11 caves in northern Spain, which suggests that Europe's earliest cave art was non-figurative in nature and that animal paintings did not appear until considerably later^{11,28}. Currently, the oldest dated rock art motif in Europe (and the world) is from El Castillo, where a single thin calcite deposit overlying a red 'disk' yielded a minimum uranium-series age of 40.8 kyr¹¹. The alternative model is that cave art first appeared in Europe in fully developed form, as implied

Table 1 | Results of uranium-series disequilibrium dating showing the minimum age of each dated rock art motif

Sample	Site	Description	$^{230}\text{Th}/^{238}\text{U}$	$^{234}\text{U}/^{238}\text{U}$	$^{230}\text{Th}/^{232}\text{Th}$	Uncorrected age (kyr)	+2 σ (kyr)	−2 σ (kyr)	Corrected age (kyr)	+2 σ (kyr)	−2 σ (kyr)	Initial $^{234}\text{U}/^{238}\text{U}$
LL3.2	Leang Lompoa	Overlies hand stencil	0.1525 ± 0.0022	1.0067 ± 0.0014	137	17.87	0.27	0.28	17.77	0.42	0.42	1.0070 ± 0.0014
LB2.3	Leang Barugayya 1	Overlies hand stencil	0.1624 ± 0.0077	0.9812 ± 0.0027	858	20.00	1.00	1.00	19.70	1.00	1.00	0.9801 ± 0.0028
LB3.3	Leang Barugayya 1	Overlies hand stencil	0.2004 ± 0.0214	0.9799 ± 0.0025	428	24.90	2.90	2.90	24.90	3.10	3.00	0.9784 ± 0.0026
GJ2.2	Gua Jing	Overlies hand stencil	0.1996 ± 0.0044	0.9943 ± 0.0009	50	24.40	0.60	0.59	24.00	1.10	1.10	0.9939 ± 0.0009
LB1.2	Leang Barugayya 1	Overlies hand stencil	0.2308 ± 0.0211	0.9831 ± 0.0025	360	29.10	3.00	2.90	29.10	3.20	3.10	0.9817 ± 0.0028
LL1.3	Leang Lompoa	Overlies hand stencil	0.2322 ± 0.0030	1.0128 ± 0.0024	121	28.31	0.44	0.43	28.10	0.66	0.67	1.0138 ± 0.0025
LL2.2	Leang Lompoa	Overlies hand stencil	0.2391 ± 0.0064	1.0065 ± 0.0007	133	29.50	0.92	0.89	29.30	1.20	1.10	1.0070 ± 0.0008
GJ1.3	Gua Jing	Sequence of aliquots	0.2525 ± 0.0048	0.9998 ± 0.0010	31	31.70	0.69	0.69	30.90	1.70	1.80	0.9998 ± 0.0011
LS1.2	Leang Sampeang	Overlies hand stencil	0.2549 ± 0.0044	0.9823 ± 0.0007	324	32.70	0.66	0.65	32.60	0.76	0.76	0.9806 ± 0.0007
LJ2	Leang Jarie	Overlies hand stencil	0.2738 ± 0.0022	0.9942 ± 0.0010	422	35.04	0.32	0.32	34.98	0.41	0.41	0.9935 ± 0.0011
LT1.2	Leang Timpuseng	Overlies babirusa painting	0.2927 ± 0.0100	1.0163 ± 0.0023	682	37.00	1.50	1.50	36.90	1.60	1.50	1.0181 ± 0.0025
LB4.2	Leang Barugayya 2	Overlies undetermined animal figure	0.3481 ± 0.0385	1.0080 ± 0.0042	18	46.00	6.40	6.20	44.00	9.10	8.30	1.0091 ± 0.0046
LJ1	Leang Jarie	Overlies hand stencil	0.3006 ± 0.0018	0.9839 ± 0.0014	1,474	39.69	0.29	0.30	39.67	0.32	0.32	0.9820 ± 0.0015
LT2.3	Leang Timpuseng	Overlies hand stencil	0.3177 ± 0.0055	1.0156 ± 0.0011	2,845	40.80	0.83	0.83	40.70	0.87	0.84	1.0175 ± 0.0013

by the great antiquity of the elaborate animal paintings from Chauvet Cave in southern France²⁹. Although the early chronology for this art is disputed³⁰, the oldest animal image from Chauvet Cave is attributed an age of $32,410 \pm 720$ radiocarbon years BP ($\sim 35,000$ calendar years BP) on the basis of ^{14}C -dating of charcoal pigment²⁹.

Our dating results from Sulawesi suggest that figurative art was already part of the cultural repertoire of the first modern human populations to reach this region more than 40 kyr ago. It is possible that rock art emerged independently at around the same time and at roughly both ends of the spatial distribution of early modern humans. An alternative scenario, however, is that cave painting was widely practised by the first *H. sapiens* to leave Africa tens of thousands of years earlier, and thus that naturalistic animal art from Leang Timpuseng and Leang Barugayya 2, as well as Chauvet Cave in France, may well have much deeper origins outside both western Europe and Sulawesi. If so, we can expect future discoveries of depictions of human hands, figurative art and other forms of image-making dating to the earliest period of the global dispersal of our species.

METHODS SUMMARY

A small segment ($\sim 100\text{--}200\text{ mm}^2$) of each coralloid was removed from the rock art panels using a battery-operated rotary tool equipped with a diamond saw blade. Each coralloid sample was sawn *in situ* so as to produce a continuous microstratigraphic profile extending from the outer surface of the coralloid through the pigment layer and into the underlying rock face. The only exceptions were Leang Jarie 1 and 2 (2012), which were sawn *in situ* but not through the pigment layer. All of the sampled coralloids comprised multiple layers of dense and non-porous calcite. The identification of a pigment layer overlain by an extensive accumulation of calcite laminations within each coralloid (except for Leang Jarie 1 and 2 (2012)) demonstrates unambiguously that the sampled speleothems formed over the motifs (see Figs 2 and 3 and Extended Data Figs 2–9). In the laboratory, the samples were micro-excavated in arbitrary ‘spits’ over the entire surface of the coralloids, creating a series of aliquots less than 1 mm thick. The pigment layer was visible across the entire length of the sample (except for Leang Jarie 1 and 2 (2012)). In total, we obtained 55 uranium-series age determinations (a further two samples failed to produce enough signal for age determination) (Table 1 and Supplementary Information). The uranium-series isotopes were measured on a ThermoFinnigan Neptune Plus Multi-Collector inductively coupled plasma mass spectrometer at the Research School of Earth Sciences, Australian National University. Calculation of ages and initial $^{234}\text{U}/^{238}\text{U}$ ratios was done with Isoplot 3.75. Corrections for detrital components were calculated

assuming a bulk Earth $^{232}\text{Th}/^{238}\text{U}$ concentration ratio of the upper crust of $3.8 \pm 50\%$ and secular equilibrium for ^{230}Th , ^{234}U and ^{238}U . In the text, minimum ages are quoted as measured age minus 2σ and maximum ages as measured age plus 2σ rounded to one decimal place.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 February; accepted 30 April 2014.

- Lewis-Williams, D. J. *The Mind in the Cave: Consciousness and the Origins Of Art* (Thames & Hudson, 2002).
- White, R. *et al.* Context and dating of Aurignacian vulvar representations from Abri Castanet, France. *Proc. Natl Acad. Sci. USA* **109**, 8450–8455 (2012).
- Conard, N. J. Palaeolithic ivory sculptures from southwestern Germany and the origins of figurative art. *Nature* **426**, 830–832 (2003).
- Dowson, T. A. & Porr, M. in *The Archaeology of Shamanism* (ed. Price, N.) 165–177 (Routledge, 2001).
- Aubert, M. A review of rock art dating in the Kimberley, Western Australia. *J. Archaeol. Sci.* **39**, 573–577 (2012).
- Brumm, A. & Moore, M. W. Symbolic revolutions and the Australian archaeological record. *Camb. Archaeol. J.* **15**, 157–175 (2005).
- Langley, M. C., Clarkson, C. & Ulm, S. From small holes to grand narratives: the impact of taphonomy and sample size on the modernity debate in Australia and New Guinea. *J. Hum. Evol.* **61**, 197–208 (2011).
- Mellars, P. Going east: new genetic and archaeological perspectives on the modern human colonization of Eurasia. *Science* **313**, 796–800 (2006).
- Roberts, R. G., Jones, R. & Smith, M. Thermoluminescence dating of a 50,000-year-old human occupation site in northern Australia. *Nature* **345**, 153–156 (1990).
- Roberts, R. G. *et al.* The human colonisation of Australia: optical dates of 53,000 and 60,000 years bracket human arrival at Deaf Adder Gorge, Northern Territory. *Quat. Sci. Rev.* **13**, 575–583 (1994).
- Pike, A. W. G. *et al.* U-series dating of Paleolithic art in 11 caves in Spain. *Science* **336**, 1409–1413 (2012).
- McDonald, R. C. Limestone morphology in South Sulawesi, Indonesia. *Z. Geomorphol.* **26** (suppl.), 79–91 (1976).
- Glover, I. C. Leang Burung 2: an Upper Palaeolithic rock shelter in south Sulawesi, Indonesia. *Mod. Quat. Res. SE Asia* **6**, 1–38 (1981).
- Glover, I. C. The effects of sink action on archaeological deposits in caves: an Indonesian example. *World Archaeol.* **10**, 302–317 (1979).
- van Heekeren, H. R. Rock-paintings and other prehistoric discoveries near Maros (South West Celebes). *Laporan Tahunan Dinas Purbakala* **1950**, 22–35 (1952).
- Bulbeck, D., Pasqua, M. & Di Lello, A. Culture history of the Toalean of South Sulawesi, Indonesia. *Asian Perspect.* **39**, 71–108 (2000).

17. Bulbeck, D., Sumantri, I. & Hiscock, P. Leang Sakapao 1, a second dated Pleistocene site from South Sulawesi, Indonesia. *Mod. Quat. Res. SE Asia* **18**, 111–128 (2004).
18. Eriawati, Y. *Lukisan di Gua-Gua Karst Maros–Pangkep, Sulawesi Selatan: Gambaran Penghuni dan Matapencahariannya* (Indonesian Ministry of Cultural Media Development, 2003).
19. O'Connor, S. Nine new painted rock art sites from East Timor in the context of the Western Pacific region. *Asian Perspect.* **42**, 96–128 (2003).
20. Simanjuntak, T. *Austronesian in Sulawesi* (Center for Prehistoric and Austronesian Studies, 2008).
21. Hill, C. A. & Forti, P. *Cave Minerals of the World* (National Speleological Society, 1997).
22. Fage, L. H. & Chazine, J. M. *Bornéo la Mémoire des Grottes* (Fage éditions, 2009).
23. Plagnes, V. *et al.* Cross dating (Th/U– ^{14}C) of calcite covering prehistoric paintings in Borneo. *Quat. Res.* **60**, 172–179 (2003).
24. Taçon, P. S. C. & Tan, N. H. in *Rock Art News of the World 4* (eds Bahn, P., Franklin, N. & Strecker, M.) 207–214 (Oxbow Books, 2012).
25. Chaloupka, G. *Journey in Time, the World's Longest Continuing Art Tradition: The 50,000-year Story of the Australian Aboriginal Rock Art of Arnhem Land* (Reed, 1993).
26. Morwood, M. J. *Visions from the Past: The Archaeology of Australian Aboriginal Art* (Allen & Unwin, 2002).
27. David, B. *et al.* How old are Australia's pictographs? A review of rock art dating. *J. Archaeol. Sci.* **40**, 3–10 (2013).
28. García-Díez, M. *et al.* Uranium series dating reveals a long sequence of rock art at Altamira Cave (Santillana del Mar, Cantabria). *J. Archaeol. Sci.* **40**, 4098–4106 (2013).
29. Valladas, H. *et al.* Bilan des datations carbone 14 effectuées sur des charbons de bois de la grotte Chauvet. *Bull. Soc. Préhist. Fr.* **102**, 109–113 (2005).
30. Combier, J. & Jouve, G. Nouvelles recherches sur l'identité culturelle et stylistique de la grotte Chauvet et sur sa datation par la méthode du ^{14}C . *L'anthropologie* **118**, 115–151 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements The fieldwork was authorized by the director of the Makassar Heritage Department (BPPP), M. Said, and the director of the National Centre for Archaeology in Jakarta (ARKENAS), B. Sulistyanto. We further acknowledge Balai Arkeologi Makassar, the Indonesian State Ministry of Research and Technology, and the Geological Survey Institute in Bandung, for facilitating the research. We thank the University of Wollongong's Deputy Vice Chancellor (Research), J. Raper, for additional project support. Field assistants included M. Andi Pampang and A. A. Oktaviana. Technical laboratory assistance involved G. Mortimer, H. Price, L. Sweetman and L. Yu., and C. Owers provided map data. We thank P. Taçon and M. W. Moore for critical feedback on the manuscript. This research was supported by grants from the Australian Research Council to M.A. (DP110102898/DE140100254) and A.B. (DP0879624/DE130101560) and the Centre for Archaeological Science (CAS), University of Wollongong.

Author Contributions A.B. and T.S. conceived the study with M.A., as part of a wider project led by M.R., E.W.S. and B.H., in collaboration with A.B., M.J.M. and G.v.d.B. M.A. and A.B. identified the samples. M.A. collected the samples and conducted the uranium-series dating with A.D. M.A. and A.B. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.A. (m.aubert@griffith.edu.au).

METHODS

Coralloid speleothems form from thin films of water precipitation on cave surfaces, resulting in concentric growth rings, and can be nodular, globular, botryoidal or coral-like in morphology²¹. When precipitated from saturated solutions, calcium carbonate usually contains small amounts of soluble uranium (^{238}U and ^{234}U), which eventually decay to ^{230}Th . The latter is essentially insoluble in cave waters and will not precipitate with the calcium carbonate. This produces disequilibrium in the decay chain where all isotopes in the series are no longer decaying at the same rate. Subsequently, ^{238}U and ^{234}U decay to ^{230}Th until secular equilibrium is reached. Because the decay rates are known, the precise measurement of these isotopes allows calculation of the age of the carbonate formation³¹.

It is also common for secondary calcium carbonate to be contaminated by detrital materials, such as wind-blown or waterborne sediments, and as such can lead to uranium-series ages that are erroneously older than the true age of the sample. This is because the detrital fraction will contribute to the overall amount of uranium-series nuclides so that the sample does not reflect a radioactive disequilibrium related to the time of carbonate formation. The effects of detrital contamination can be identified and often corrected for by measuring the activity of ^{232}Th that is solely present in the detrital fraction but which plays no part in the decay chain of uranium. An indication of the degree of detrital contamination is expressed as $^{230}\text{Th}/^{232}\text{Th}$ activity, with high values (>20) indicating little or no effect on the calculated age and low values (<20) indicating that the correction on the age will be significant³¹. Except for two samples (LL3.1 and B4.2), all our samples have $^{230}\text{Th}/^{232}\text{Th}$ activity >20 , indicating sample purity.

Sample preparation was conducted at the Wollongong Isotope Geochronology Laboratory, University of Wollongong. The small calcium carbonate samples were weighed separately in Savillex perfluoroalkoxy polymer resin (PFA) vials. The samples were covered with MilliQ water, and drops of Merck Ultrapur 60% HNO_3 were added until complete dissolution was achieved. A spike solution enriched in ^{236}U – ^{229}Th was subsequently added and the mixture was left to equilibrate overnight. The solutions were evaporated to dryness and then redissolved in 1.5 M HNO_3 ready for ion-exchange chromatography, consisting of 0.25 ml of Eichrom TRU resin over 0.1 ml of Eichrom pre-filter resin. The resins were cleaned by passing 3 M HCl, 0.2 M HCl and a 0.1 M HCl + 0.3 M HF mixture through the columns before use and then preconditioned with 1.5 M HNO_3 . After the sample solutions had been loaded on the TRU resin bed as solutions in 1.5 M HNO_3 , the columns were washed with 1.5 M HNO_3 and 3 M HCl. Uranium and thorium were imperfectly separated from the ion-exchange medium with 0.2 M HCl (for thorium), and 0.1 M HCl + 0.3 M HF (for uranium). Finally, the samples were evaporated to dryness and redissolved in 4 ml of 2% HNO_3 .

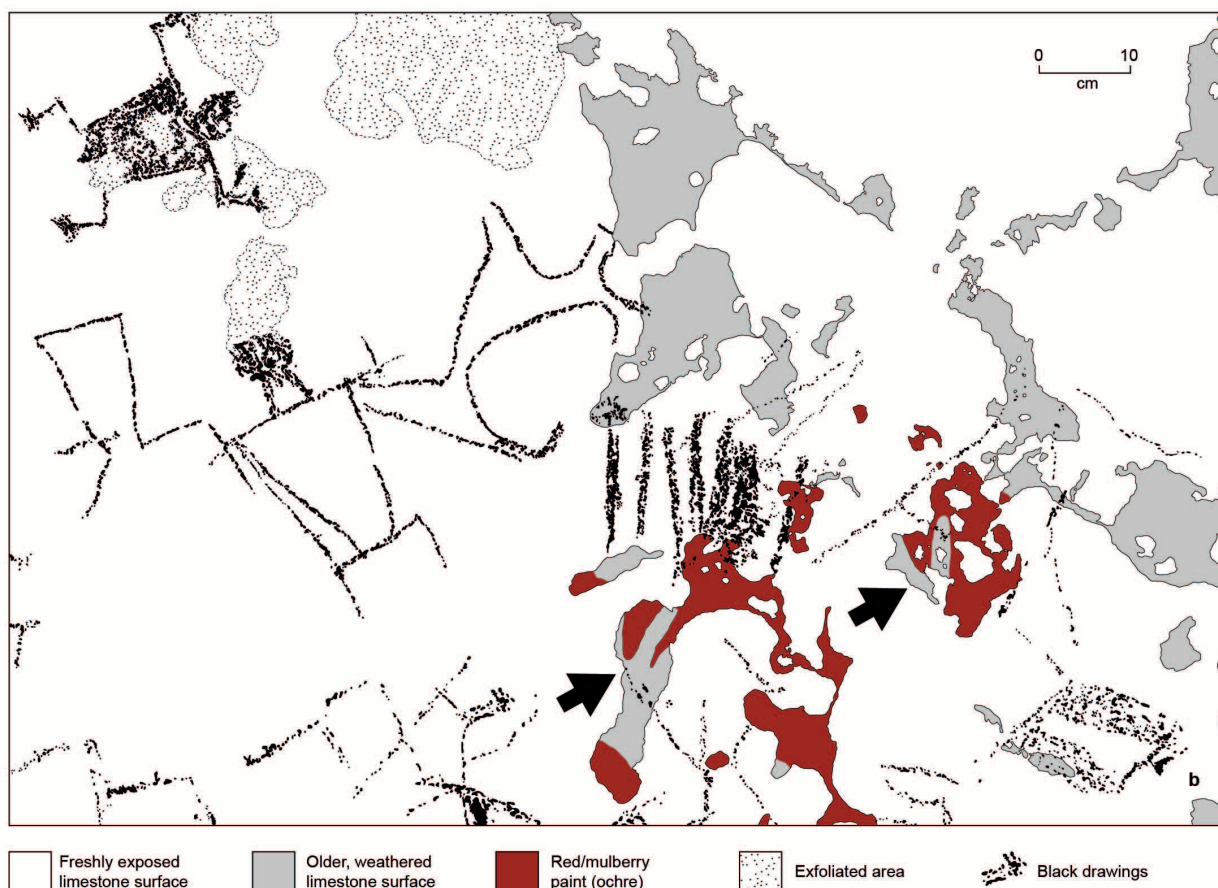
The U and Th solutions were introduced separately into a ThermoFinnigan Neptune Plus Multi-Collector inductively coupled plasma mass spectrometer at the

Research School of Earth Sciences, Australian National University. The Neptune Plus is equipped with a large interface pump, Jet Sample and Skimmer cones, electrostatic analyser, secondary electron multiplier (SEM) and retarding potential quadrupole (RPQ) for high abundance sensitivity. Samples were aspirated using an electrospray ionization PFA-ST Aridus II nebulizer at an uptake rate of $\sim 0.1 \text{ ml min}^{-1}$. The sweep gas (Ar) flow rate was set to $\sim 3\text{--}4 \text{ l min}^{-1}$ and nitrogen was set to $\sim 2\text{--}4 \text{ ml min}^{-1}$. Sensitivity was $>1 \text{ V per p.p.b. U}$.

Uranium isotopes were measured with the RPQ off; thorium isotopes were measured with the RPQ on. Isotopic ratios were corrected for background, tailing of ^{238}U on ^{236}U and ^{234}U , SEM/Faraday yield and instrumental mass bias (using $^{238}\text{U}/^{235}\text{U} = 137.88$) after subtraction of the minor spike component. The SEM/Faraday yield was calculated externally with the NBS 960 standard by alternating ^{235}U between the SEM and Faraday array while measuring ^{238}U on the Faraday array. This value was corrected for instrumental mass bias and compared with the true value in SRM 960 = 0.007265. The SRM 960 standard was measured every two samples. Relative gains derived from standard measurements were then interpolated to the unknowns. Other standards used in this study were AC-1, an Australian National University (ANU) coral powder with a measured TIMS U-series age of 125,550 years³², and HU-1, a solution of secular equilibrium Harwell Uraninite, also supplied by the ANU. AC-1 and HU-1 results are shown in Supplementary Information, and in both cases are within the error of the expected values. Total procedure blanks were in the order of 0.9 pg for Th and 0.1 pg for U. Further details on our multi-collector inductively coupled plasma mass spectrometry procedure can be found in ref. 33.

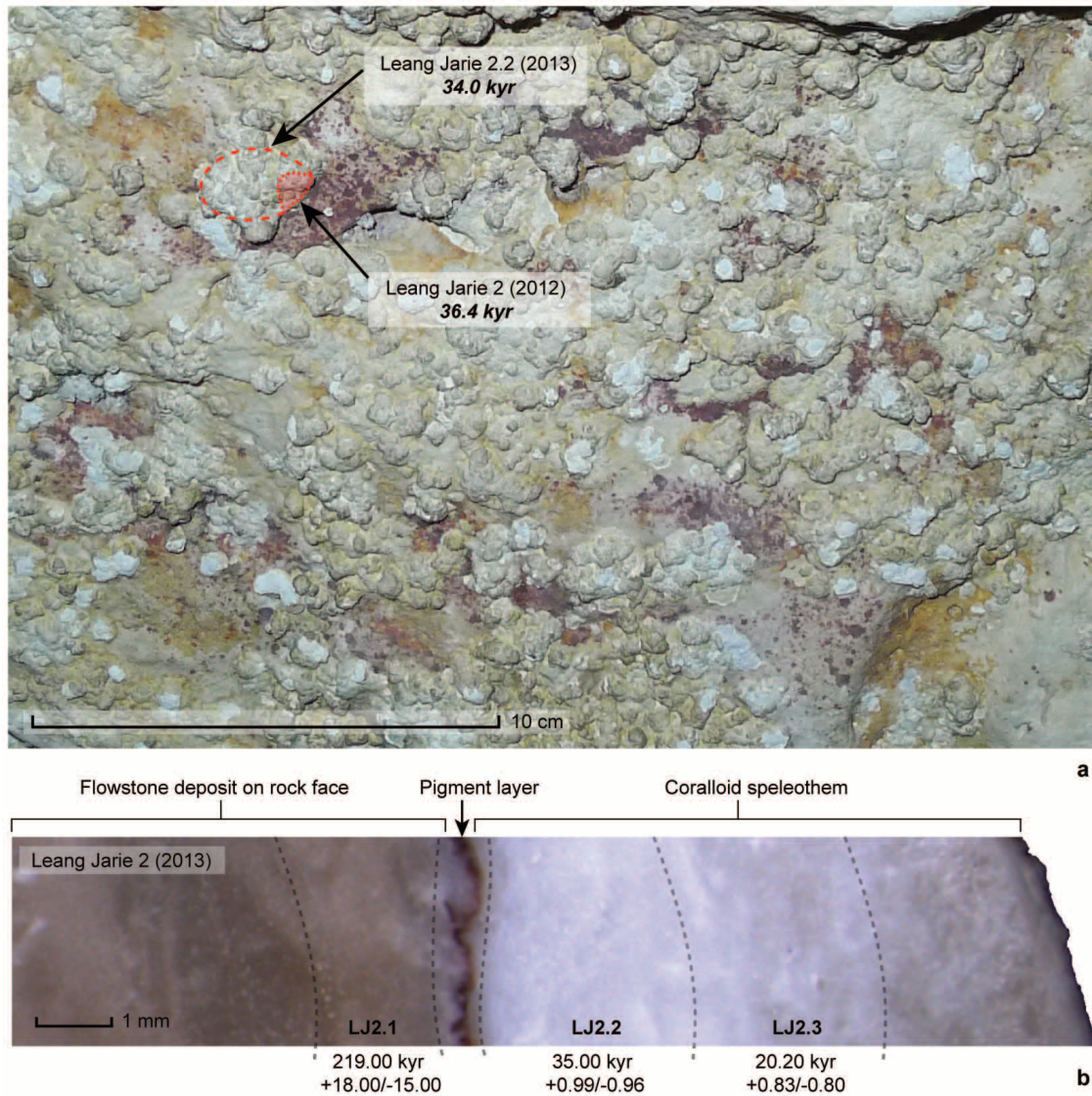
Calculation of ages and initial $^{234}\text{U}/^{238}\text{U}$ ratios was performed with Isoplot 3.75 using the following decay constants (dc) and half-lives (hl): $^{238}\text{U}_{\text{dc}} = 1.55125 \times 10^{-10}$; $^{238}\text{U}_{\text{hl}} = 4.46831 \times 10^9$; $^{234}\text{U}_{\text{dc}} = 2.82207 \times 10^{-6}$; $^{234}\text{U}_{\text{hl}} = 2.45617 \times 10^5$; $^{232}\text{Th}_{\text{dc}} = 4.94752 \times 10^{-11}$; $^{232}\text{Th}_{\text{hl}} = 1.401 \times 10^{10}$; $^{230}\text{Th}_{\text{dc}} = 9.17052 \times 10^{-6}$; $^{230}\text{Th}_{\text{hl}} = 7.55843 \times 10^4$. Errors were calculated by Monte Carlo simulation (5,000 trials), ignoring the uncertainties in the ^{235}U and ^{238}U decay constants. Corrections for detrital components were calculated assuming a bulk Earth $^{232}\text{Th}/^{238}\text{U}$ concentration ratio of the upper crust of $3.8 \pm 50\%$ ³⁴ and secular equilibrium for ^{230}Th , ^{234}U and ^{238}U .

1. Bourdon, B., Henderson, G. M., Lundstrom, C. C. & Turner, S. P. *Uranium-series Geochemistry* (Mineralogical Society of America, 2003).
2. McCulloch, M. T. & Esat, T. The coral record of last interglacial sea levels and sea surface temperatures. *Chem. Geol.* **169**, 107–129 (2000).
3. McCulloch, M. T. & Mortimer, G. E. Applications of the ^{238}U – ^{230}Th decay series to dating of fossil and modern corals using MC-ICPMS. *Aust. J. Earth Sci.* **55**, 955–965 (2008).
4. Cheng, H., Adkins, J., Edwards, R. & Boyle, E. U–Th dating of deep-sea corals. *Geochim. Cosmochim. Acta* **64**, 2401–2416 (2000).
5. Leclerc, P. in *Expédition Thai-Maros* 86 147–153 (Association Pyrénéenne de Spéléologie, 1987).



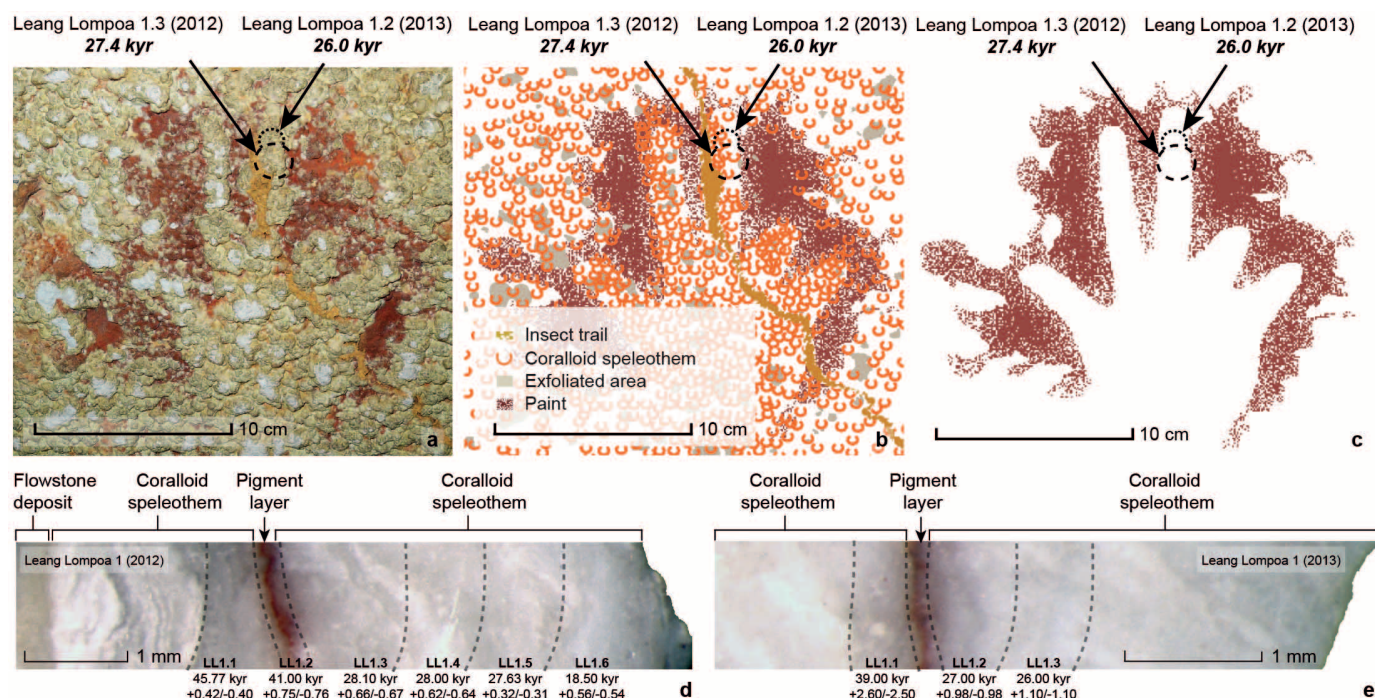
Extended Data Figure 1 | Rock art panel on the ceiling at Leang Bulu Bettue.
a, Black drawings of early Austronesian style were made on a relatively freshly exposed limestone surface and are superimposed over remnant patches of a much older surface, now extremely heavily weathered and almost completely

exfoliated, containing faded hand stencils (shown more clearly and highlighted by arrows in **b**). The same rock art panel was documented and illustrated in a publication by a team of French cavers in 1986, but the hand stencils were not identified³⁵.



Extended Data Figure 2 | Dated rock art from Leang Jarie. **a**, Locations of the sampled coraloid speleothems and associated hand stencils. **b**, Profile of the coraloid speleothem showing the microexcavated subsamples bracketing the

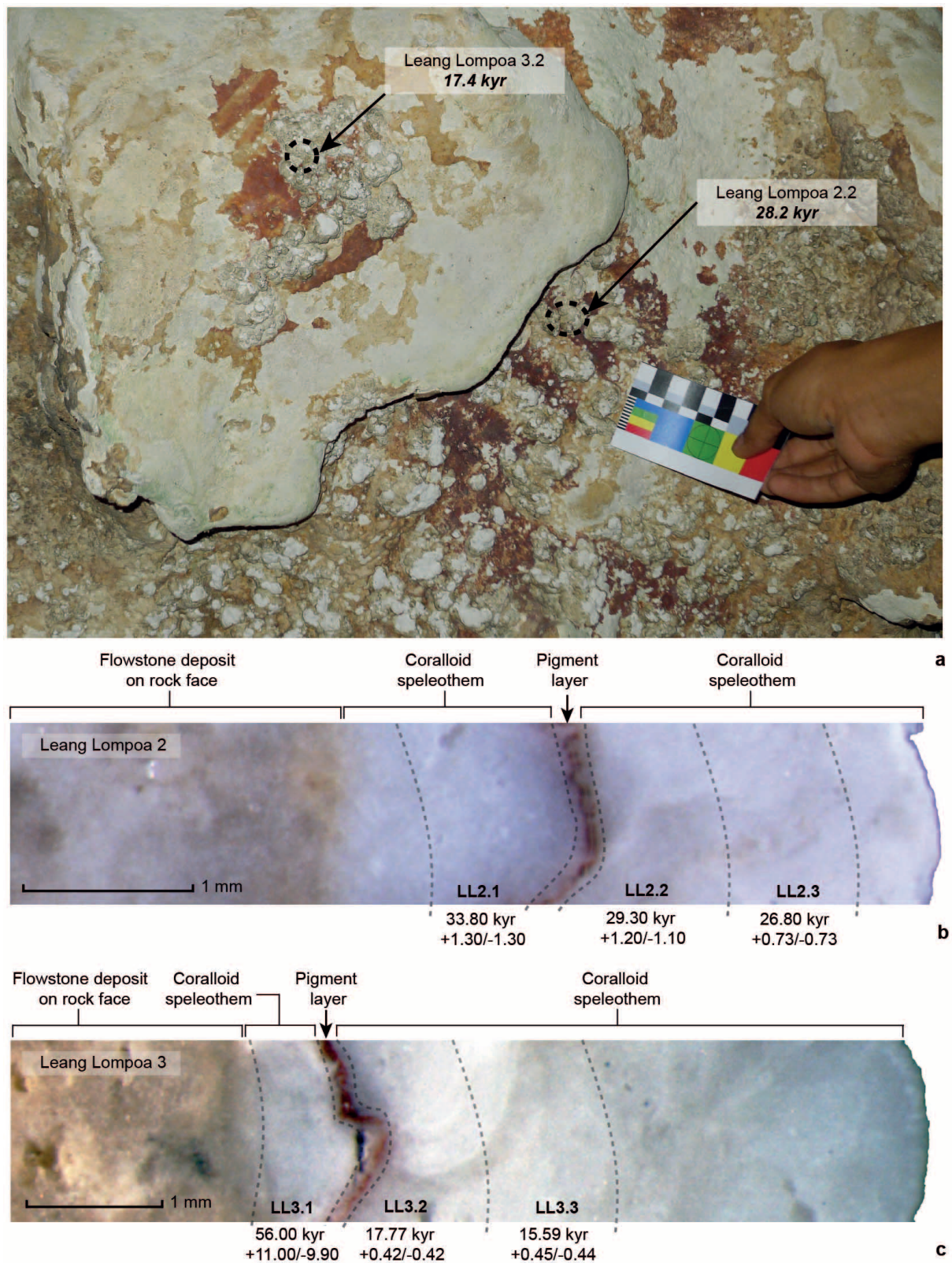
age of the paintings. The Leang Jarie 2 (2012) sample is from above the pigment layer and so only provides a minimum age for the underlying hand stencils.



Extended Data Figure 3 | Dated rock art from Leang Lompoa.

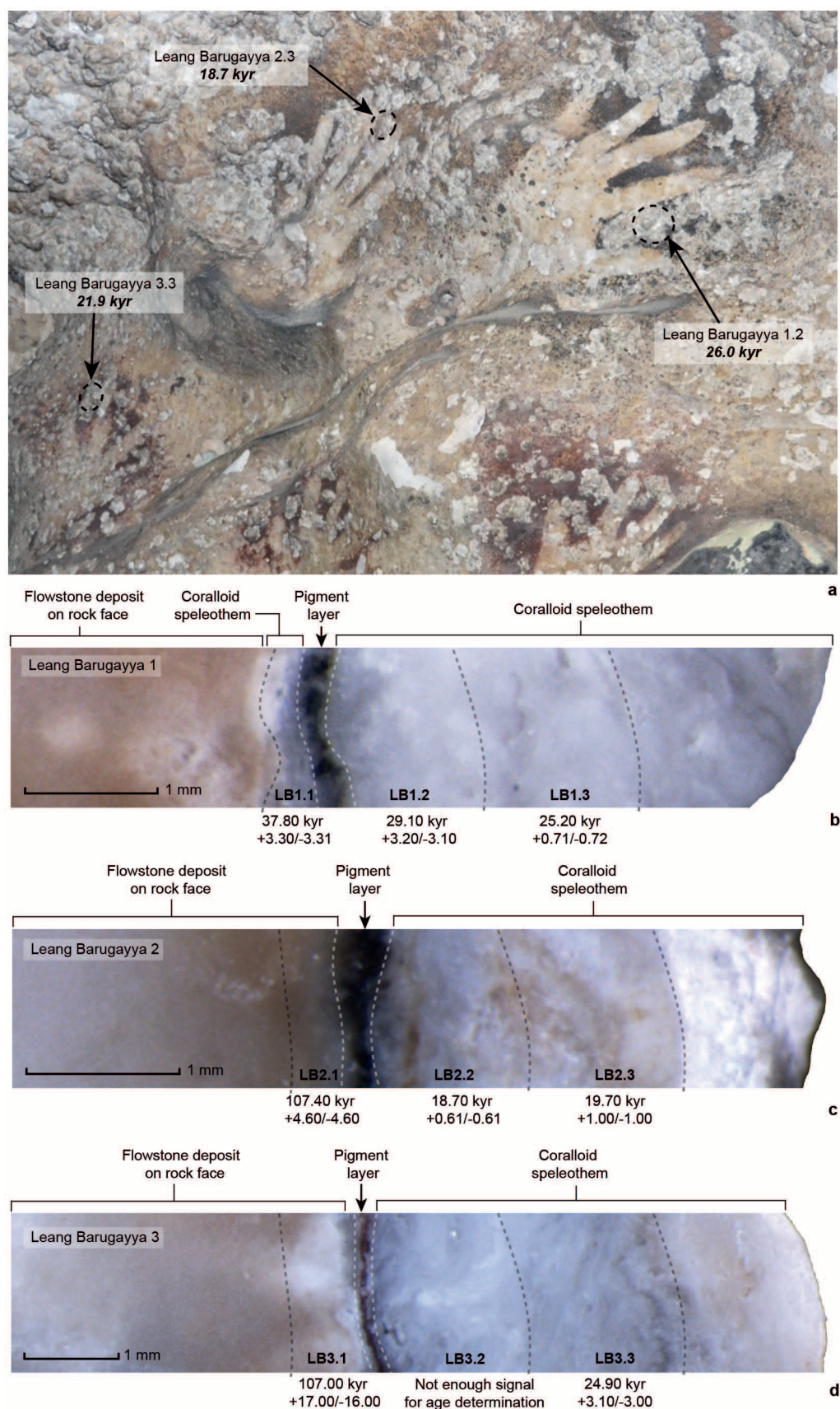
a, Photograph showing the locations of the sampled coralloid speleothems and associated hand stencil. **b**, **c**, Tracings showing the locations of the sampled coralloid speleothems and associated hand stencil. Although heavily obscured by coralloid speleothems, we interpret this image as a 'mutilated hand' stencil, which shows in outline a human hand with two amputated digits or with the third and fourth fingers folded into the palm. The hand stencil is located on the ceiling of a narrow, dimly lit passage leading off from the main entrance to

the cave. Samples Leang Lompoa 1 (2012) and Leang Lompoa 1 (2013) are part of the same cluster of coralloid speleothems that formed over the hand stencil. **d**, **e**, Profiles of the coralloid speleothems showing the microexcavated subsamples bracketing the age of the motif. Note that sample LL1.2 (2012) does not represent the age of the hand stencil. The resultant age reflects a mixture of calcium carbonate from below and above the pigment layer. Tracing credit: Leslie Refine 'Graph & Co' (France).



Extended Data Figure 4 | Dated rock art from Leang Lompoa. **a**, Locations of the sampled coralloid speleothems and associated hand stencils. The hand stencils occur on a 2.5-m-high ceiling in a small, dimly lit side chamber leading off from the cave mouth. The stencil at the left (Leang Lompoa 3) is stylistically

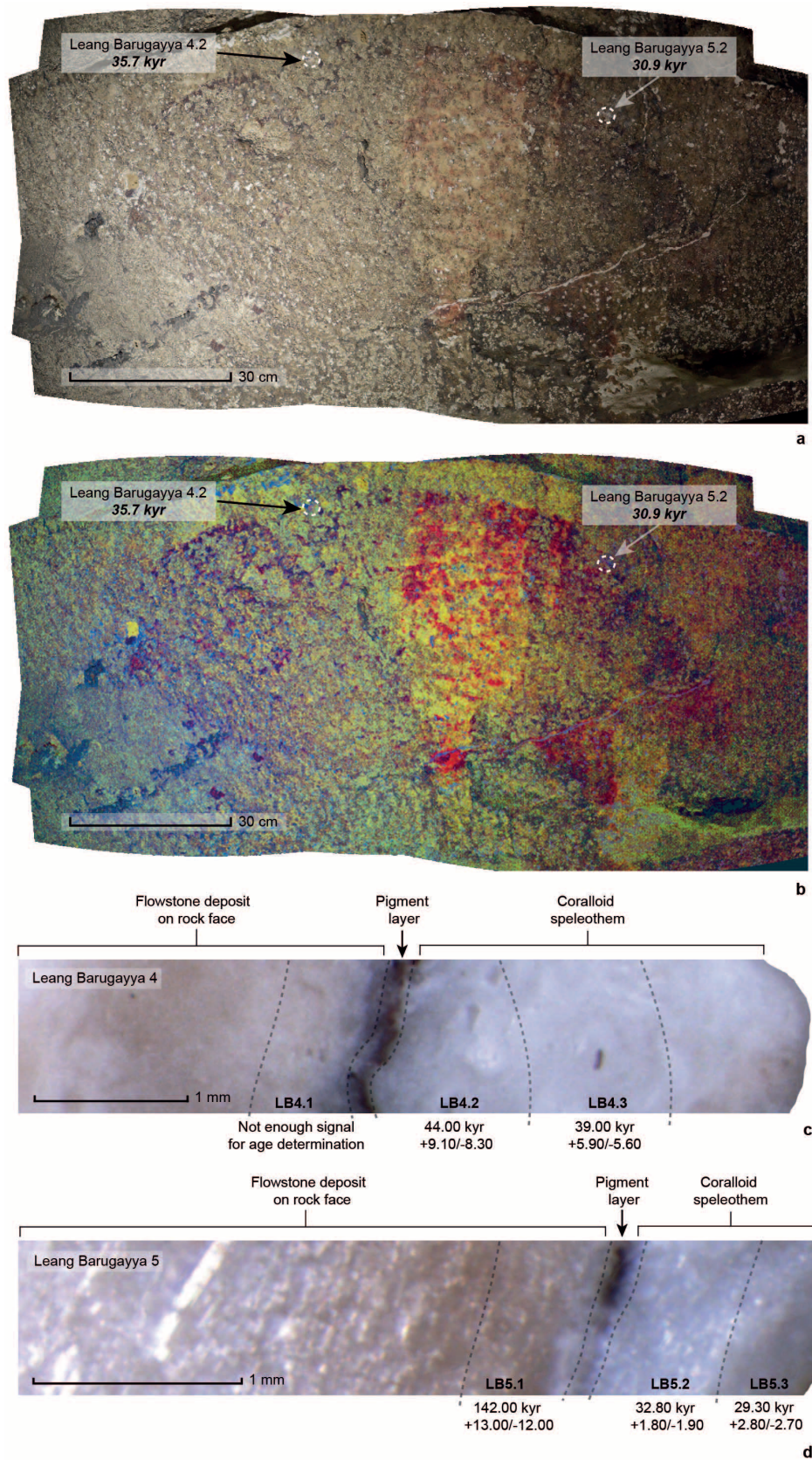
distinct from the adjacent stencil (Leang Lompoa 2), with the fingers modified by brushwork subsequent to stencilling to produce slender and pointy forms. **b, c**, Profiles of the coralloid speleothems showing the microexcavated subsamples bracketing the age of the hand stencils.



Extended Data Figure 5 | Dated rock art from Leang Barugayya 1.

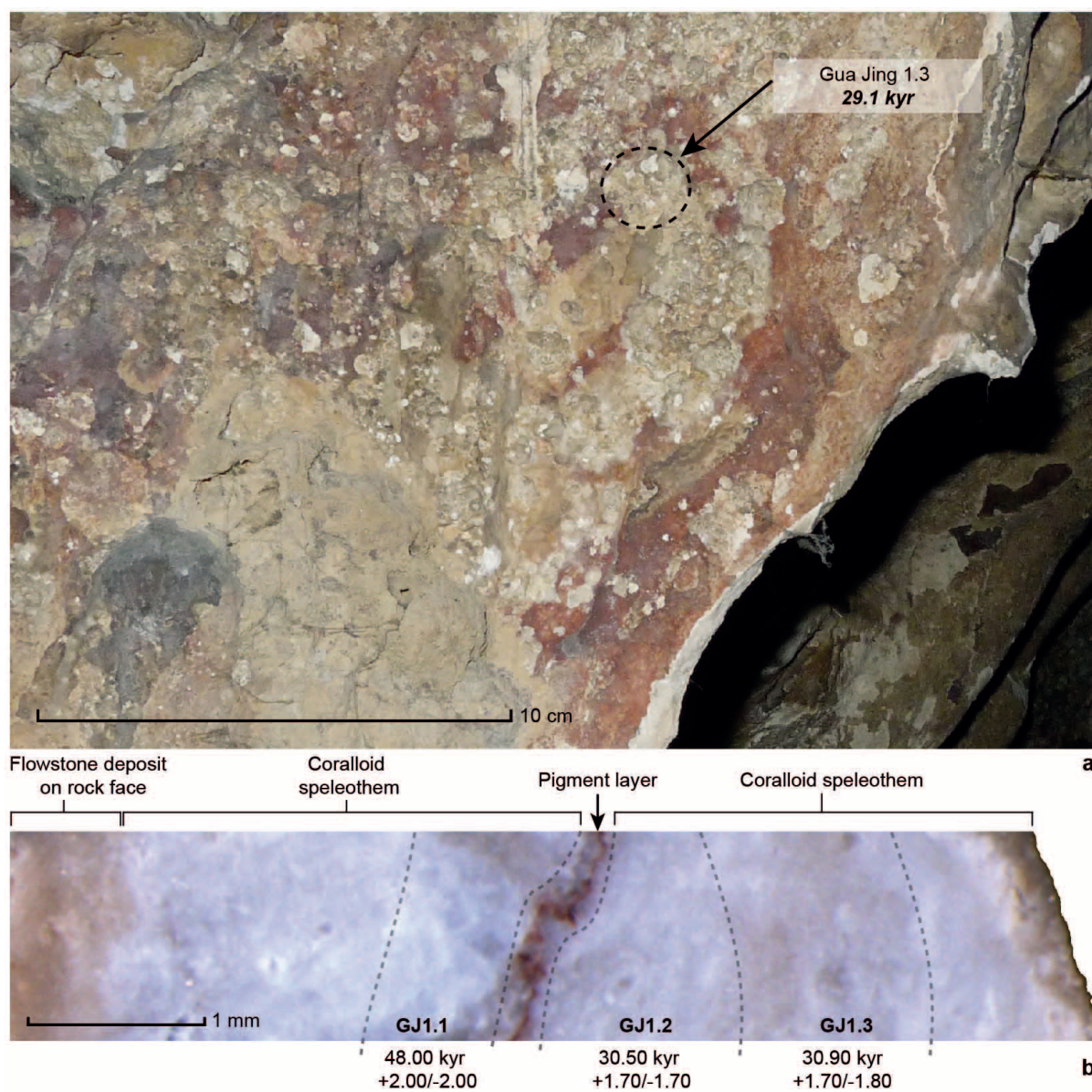
a, Locations of the sampled coralloid speleothems and associated cluster of hand stencils. The hand stencils are situated on a small rock art panel near the ceiling and close to the cave entrance. Samples LB1 and LB2 come from two

distinct hand stencils that are dark mulberry (almost black) in colour. Sample LB3 is from over an adjacent red hand stencil. **b–d**, Profiles of the coralloid speleothems showing the microexcavated subsamples bracketing the age of the hand stencils.



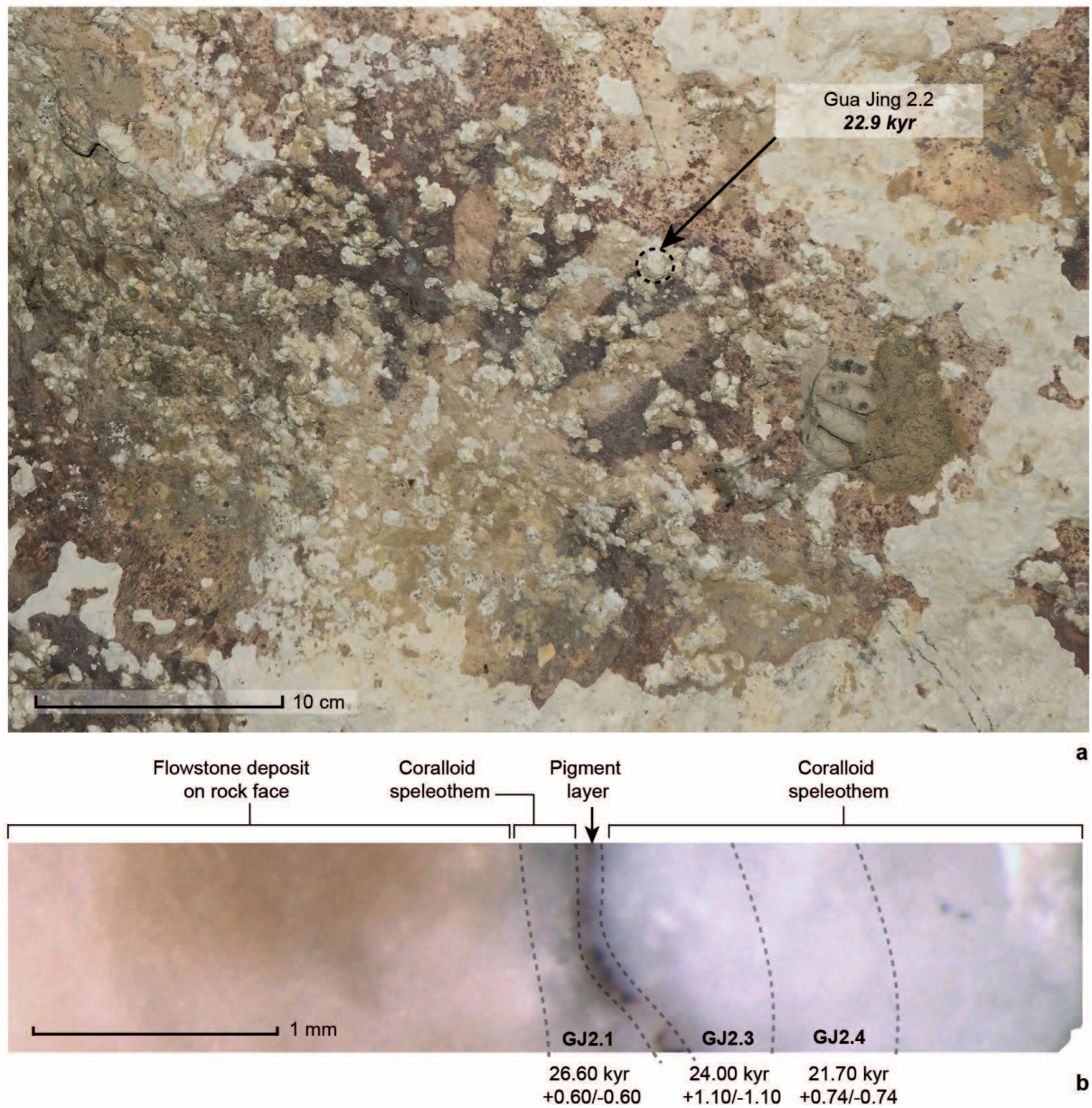
Extended Data Figure 6 | Dated animal painting from Leang Barugayya 2. **a, b**, Composite of photographs showing the locations of the sampled coraloid speleothems and associated large infilled red painting of an animal. Field photographs were altered in the software program DStretch to enhance the image (**b**). The animal species depicted is unidentified as a result of the extent of weathering and deterioration of the painting and the thick accumulation of

coralloids over the art; however, the painting seems to show in profile a large land mammal, probably a pig (a babirusa or *Sus celebensis*), with the head facing right and the hindquarters at the left. **c, d**, Profile of the coraloid speleothems showing the microexcavated subsamples bracketing the age of the painting. Images **a** and **b** courtesy of A. A. Oktaviana.

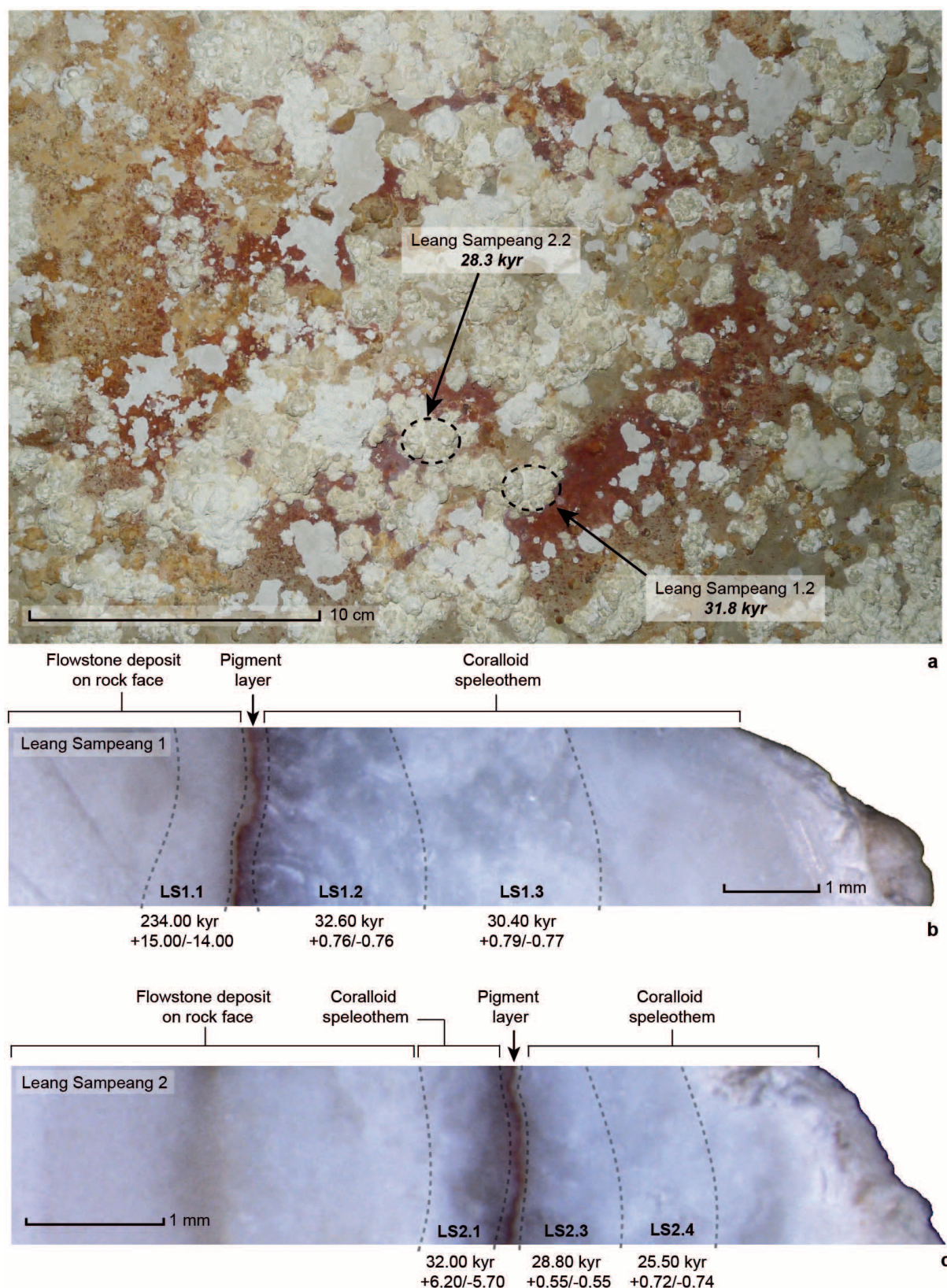


Extended Data Figure 7 | Dated rock art from Gua Jing. **a**, Location of the sampled coraloid speleothem and associated hand stencil. The hand stencil is located on a stalactite curtain 15 m from the cave entrance and 2 m above the current cave floor. The cave itself comprises a dark, winding phreatic tube

containing an extensive gallery of hand stencils and figurative animal motifs. **b**, Profile of the coraloid speleothem showing the microexcavated subsamples bracketing the age of the hand stencil.



Extended Data Figure 8 | Dated rock art from Gua Jing. **a**, Location of the sampled coralloid speleothem and associated hand stencil. **b**, Profile of the coralloid speleothem showing the microexcavated subsamples bracketing the age of the hand stencil.



Extended Data Figure 9 | Dated rock art from Leang Sampeang.

a, Locations of the sampled coralloid speleothems and associated hand stencil. Leang Sampeang is a 20-m-deep, narrow chamber with paintings located on the ceiling at the back of the cave in complete darkness. In this area the cave is only 2.5 m wide and requires crawling to reach. Samples Leang Sampeang 1 and

Leang Sampeang 2 came from the same red hand stencil located 17 m from the cave entrance and 18 cm above the current cave floor. **b**, **c**, Profiles of the coralloid speleothems showing the microexcavated subsamples bracketing the age of the hand stencil.

Inappropriate p53 activation during development induces features of CHARGE syndrome

Jeanine L. Van Nostrand¹, Colleen A. Brady^{1†}, Heiyoun Jung¹, Daniel R. Fuentes², Margaret M. Kozak^{1†}, Thomas M. Johnson^{1†}, Chieh-Yu Lin³, Chien-Jung Lin⁴, Donald L. Swiderski⁵, Hannes Vogel³, Jonathan A. Bernstein⁶, Tania Attié-Bitach^{7,8}, Ching-Pin Chang⁹, Joanna Wysocka^{2,4}, Donna M. Martin^{10,11} & Laura D. Attardi^{1,12}

CHARGE syndrome is a multiple anomaly disorder in which patients present with a variety of phenotypes, including ocular coloboma, heart defects, choanal atresia, retarded growth and development, genitourinary hypoplasia and ear abnormalities¹. Despite 70–90% of CHARGE syndrome cases resulting from mutations in the gene *CHD7*, which encodes an ATP-dependent chromatin remodeller, the pathways underlying the diverse phenotypes remain poorly understood². Surprisingly, our studies of a knock-in mutant mouse strain that expresses a stabilized and transcriptionally dead variant of the tumour-suppressor protein p53 ($p53^{25,26,53,54/3}$), along with a wild-type allele of p53 (also known as *Trp53*), revealed late-gestational embryonic lethality associated with a host of phenotypes that are characteristic of CHARGE syndrome, including coloboma, inner and outer ear malformations, heart outflow tract defects and craniofacial defects. We found that the $p53^{25,26,53,54}$ mutant protein stabilized and hyperactivated wild-type p53, which then inappropriately induced its target genes and triggered cell-cycle arrest or apoptosis during development. Importantly, these phenotypes were only observed with a wild-type p53 allele, as $p53^{25,26,53,54/-}$ embryos were fully viable. Furthermore, we found that *CHD7* can bind to the p53 promoter, thereby negatively regulating p53 expression, and that *CHD7* loss in mouse neural crest cells or samples from patients with CHARGE syndrome results in p53 activation. Strikingly, we found that p53 heterozygosity partially rescued the phenotypes in *Chd7*-null mouse embryos, demonstrating that p53 contributes to the phenotypes that result from *CHD7* loss. Thus, inappropriate p53 activation during development can promote CHARGE phenotypes, supporting the idea that p53 has a critical role in developmental syndromes and providing important insight into the mechanisms underlying CHARGE syndrome.

Unrestrained p53 activity induced by loss of the negative regulators MDM2 or MDM4 causes early embryonic lethality⁴. To explore the role of transcriptional activation by p53 in promoting developmental failure, we examined embryonic development in knock-in mice carrying mutations in one or both of p53's two transcriptional activation domains ($p53^{25,26}$ and $p53^{25,26,53,54}$; Fig. 1a and Extended Data Fig. 1). These mutations both disrupt the p53–MDM2 interaction—recapitulating loss of MDM2 regulation—and compromise transactivation³. Expression of $p53^{25,26}$, whose ability to transactivate most (but not all) p53 target genes is severely impaired, caused early embryonic lethality⁵ (at embryonic day (E) 10.5; Extended Data Fig. 2a–c). Surprisingly, $p53^{25,26,53,54/+}$ embryos ($p53^{25,26,53,54}$ is transactivation dead) also exhibited embryonic lethality but between E13.5 and E15.5 (Extended Data Tables 1 and 2). Lethality associated with $p53^{25,26,53,54}$ depended on the presence of a

wild-type p53 allele, as $p53^{25,26,53,54/-}$ adult mice were viable (Extended Data Fig. 2b, d). This observation is in contrast to the embryonic lethality of $p53^{25,26/-}$ mice, which probably results from the residual transactivation potential of p53^{25,26} on genes such as *Bax*³ (Extended Data Fig. 2a, b). Our findings underscore the importance of transactivation for the embryonic lethality induced by stabilized p53 (because $p53^{25,26,53,54/-}$ mice are viable), as well as revealing an intriguing genetic interaction between $p53^{25,26,53,54}$ (which is transactivation dead) and wild-type p53 during development.

Analysis of $p53^{25,26,53,54/+}$ embryos identified numerous gender-independent developmental phenotypes that were absent in littermate controls (Figs 1 and 2 and Extended Data Figs 3 and 4). E13.5 and older $p53^{25,26,53,54/+}$ embryos commonly displayed exencephaly, as well as craniofacial defects, including square-shaped faces, short lower jaws, cleft lip and cleft palate (Fig. 1b, c and Extended Data Fig. 4a, b). Furthermore, $p53^{25,26,53,54/+}$ embryos displayed defects in external ear formation (Fig. 1d) and a spectrum of inner ear defects, ranging from mild (for example, the posterior semi-circular canal (SCC) was either truncated or fused to the common crus (Fig. 1e)) to highly abnormal (for example, extreme inner ear bone malformation). We also observed retinal coloboma in $p53^{25,26,53,54/+}$ embryos (Fig. 1f, g). As a potential contributor to the craniofacial defects, we examined osteogenesis⁶ and found delayed bone formation in $p53^{25,26,53,54/+}$ embryos, suggesting growth retardation (Extended Data Fig. 4c). Notably, this constellation of phenotypes is reminiscent of those in human CHARGE syndrome. In particular, the combined presentation of coloboma and inner ear defects is characteristic of CHARGE syndrome and rarely occurs in other conditions⁷.

Given this phenotypic overlap with CHARGE syndrome, we examined whether $p53^{25,26,53,54/+}$ embryos display other CHARGE-associated characteristics^{8–10}. The hearts in $p53^{25,26,53,54/+}$ embryos possessed the full complement of cell types (Extended Data Fig. 4d) but displayed outflow tract defects (for example, persistent truncus arteriosus and double outlet right ventricle; Fig. 2a and Extended Data Fig. 4e, f) accompanied by ventricular septation defects (data not shown). The atrioventricular cushions also failed to undergo remodelling, foreshadowing potential heart valve defects (Fig. 2b and Extended Data Fig. 4f). Notably, outflow tract and atrioventricular septation defects are highly overrepresented in patients with CHARGE syndrome compared with individuals with isolated congenital heart disease¹¹. Additionally, the kidneys and thymus in $p53^{25,26,53,54/+}$ embryos were smaller than those in controls, and the kidneys displayed branching defects (Fig. 2c and Extended Data Fig. 4g). These phenotypes were in contrast to the liver, which exhibited normal

¹Department of Radiation Oncology, Division of Radiation and Cancer Biology, Stanford University School of Medicine, Stanford, California 94305, USA. ²Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, California 94305, USA. ³Department of Pathology, Stanford University School of Medicine, Stanford, California 94305, USA. ⁴Department of Developmental Biology, Stanford University School of Medicine, Stanford, California 94305, USA. ⁵Department of Otolaryngology, The University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. ⁶Department of Pediatrics, Stanford University School of Medicine, Stanford, California 94305, USA. ⁷Département de Génétique, Hôpital Necker-Enfants Malades, APHP, 75015 Paris, France. ⁸Unité INSERM U1163, Université Paris Descartes-Sorbonne Paris Cité, Institut Imagine, 75015 Paris, France. ⁹Krannert Institute of Cardiology, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA. ¹⁰Department of Pediatrics, The University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. ¹¹Department of Human Genetics, The University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. ¹²Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA. †Present addresses: Cardiovascular Research Center and Division of Cardiology, Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Charlestown, Massachusetts 02129, USA (C.A.B.); Department of Medicine, University of Central Florida, Orlando, Florida 32827, USA (M.M.K.); Department of Emergency Medicine, Oregon Health and Science University, Portland, Oregon 97239, USA (T.M.J.).

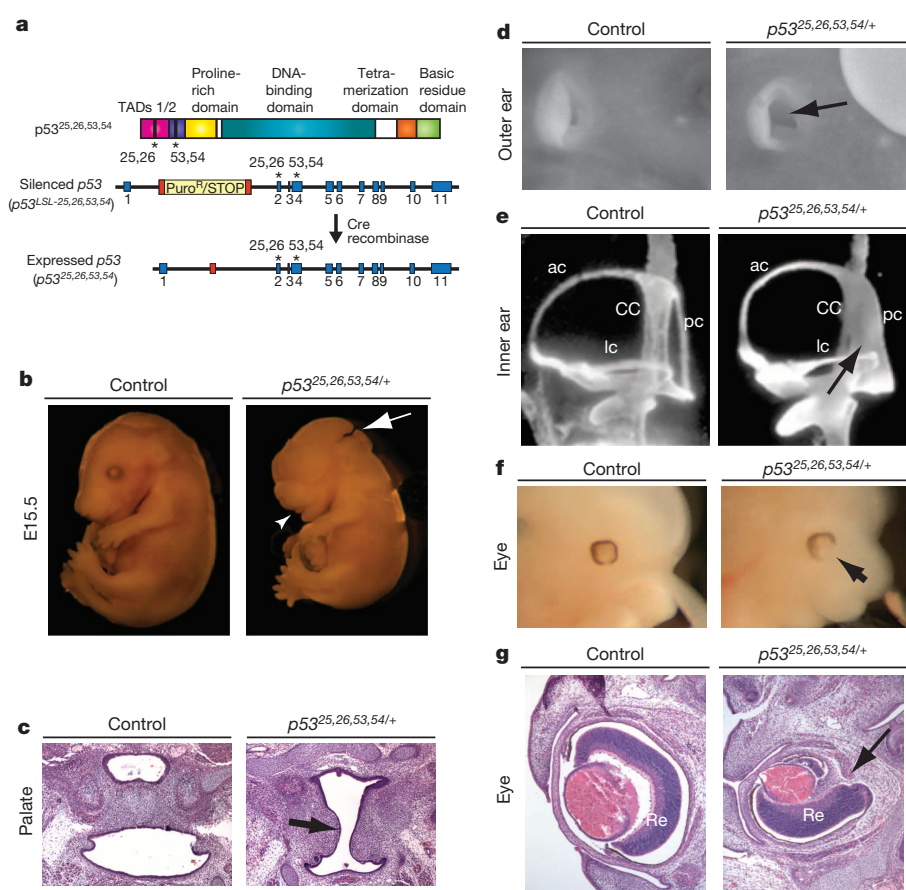


Figure 1 | $p53^{25,26,53,54/+}$ embryos exhibit lethality and diverse craniofacial defects characteristic of CHARGE syndrome. **a**, Top, p53 transcriptional activation domain (TAD) mutant protein with Leu25Gln, Trp26Ser, Phe53Gln, Phe54Ser mutations. Centre and bottom, Cre deletes the *lox-Stop-lox* (LSL) cassette, inducing $p53^{25,26,53,54/+}$ allele expression. Numbers in p53 locus schematic indicate exons. Asterisks denote mutations. Puro^R, puromycin resistance. **b**, Exencephaly (63%, $n = 35$; arrow) and short lower jaw (74%, $n = 27$; arrowhead) in an E15.5 $p53^{25,26,53,54/+}$ embryo. Original magnification, $\times 6.3$. **c**, Cleft palate (arrow) in an E15.5 $p53^{25,26,53,54/+}$ embryo ($n = 3$). Original magnification, $\times 100$. **d**, Absent external ear pinna (arrow) in an E15.5 $p53^{25,26,53,54/+}$ embryo. (47%, $n = 17$). Original magnification, $\times 40$. **e**, Posterior semi-circular canal (pc) fused to the common crus (CC) (71%, $n = 12$; arrow) in the inner ear of an E13.5 $p53^{25,26,53,54/+}$ embryo. ac, anterior canal; lc, lateral canal. Original magnification, $\times 20$. **f**, Coloboma (arrow) in an E13.5 $p53^{25,26,53,54/+}$ embryo (59%, $n = 17$). Original magnification, $\times 20$. **g**, Retinal coloboma (arrow) in an E15.5 $p53^{25,26,53,54/+}$ embryo. Re, retina. Original magnification, $\times 100$.

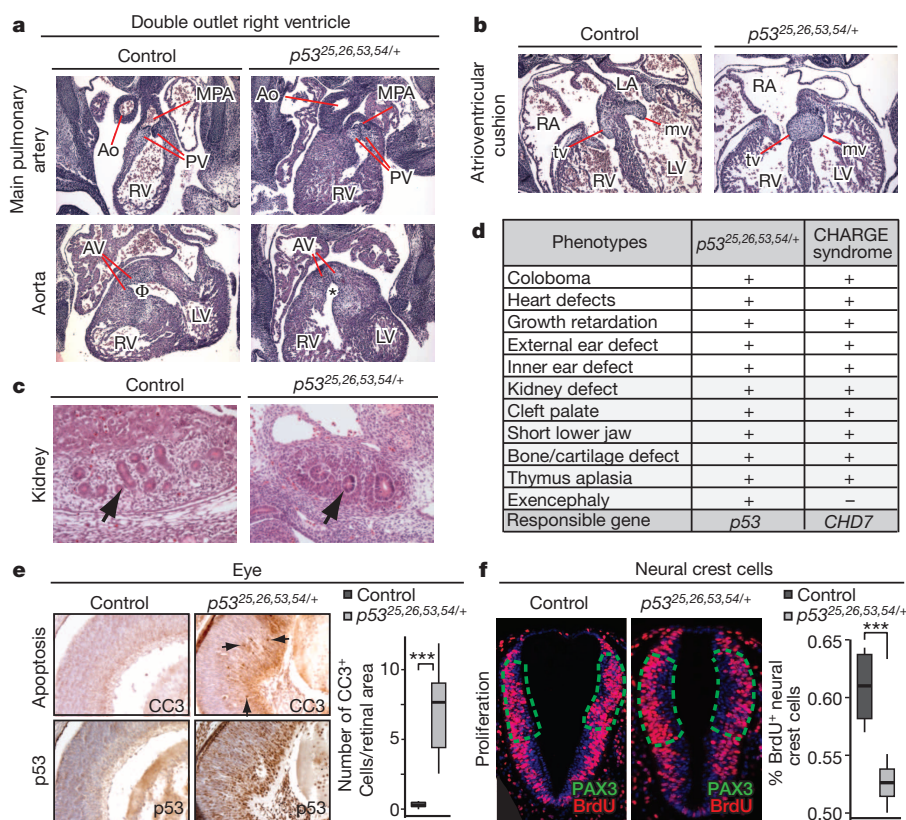


Figure 2 | $p53^{25,26,53,54/+}$ embryos exhibit additional features of CHARGE syndrome and p53-dependent cellular responses. **a**, Double outlet right ventricle (DORV) in an E13.5 $p53^{25,26,53,54/+}$ heart (50%, $n = 6$). Top, The main pulmonary artery (MPA) connects via the pulmonary valve (PV) to the right ventricle (RV) in both the control and $p53^{25,26,53,54/+}$ embryos. Bottom, The aorta (Ao) in the control embryo connects to the left ventricle (LV) via the aortic valve (AV) (Φ). The aorta in the $p53^{25,26,53,54/+}$ embryo connects to the RV via the AV (*). Original magnification, $\times 100$. **b**, Abnormal atrioventricular cushions in an E13.5 $p53^{25,26,53,54/+}$ heart (75%, $n = 4$) fail to elongate into mature mitral (mv, arrowhead) and tricuspid (tv, arrow) valves. LA, left atrium; RA, right atrium. Original magnification, $\times 100$. **c**, E13.5 $p53^{25,26,53,54/+}$ kidneys are smaller (79%), with fewer glomeruli (arrow) on average than controls (13 glomeruli on average in controls versus 3 glomeruli on average in mutants; $n = 5$). Original magnification, $\times 200$. **d**, $p53^{25,26,53,54/+}$ embryonic phenotypes observed in CHARGE syndrome (+, present; -, absent). **e**, Left, Cleaved caspase 3 (CC3, top) and p53 (bottom) immunohistochemistry in E15.5 retinas. Arrows denote CC3-positive cells. Right, The number of CC3-positive cells per retinal area is shown. ***, P value = 0.007, one-tailed Welch's t -test ($n = 5$). Original magnification, $\times 400$. **f**, 5-Bromodeoxyuridine (BrdU) immunofluorescence in E9.5 PAX3-positive neural crest cells (NCCs) (delineated by the green dotted line) (see Extended Data Fig. 6c). Right, The percentage of BrdU-positive cells in the total PAX3⁺ NCC population ***, P value = 0.004, one-tailed Student's t -test ($n = 4$). Original magnification, $\times 200$.

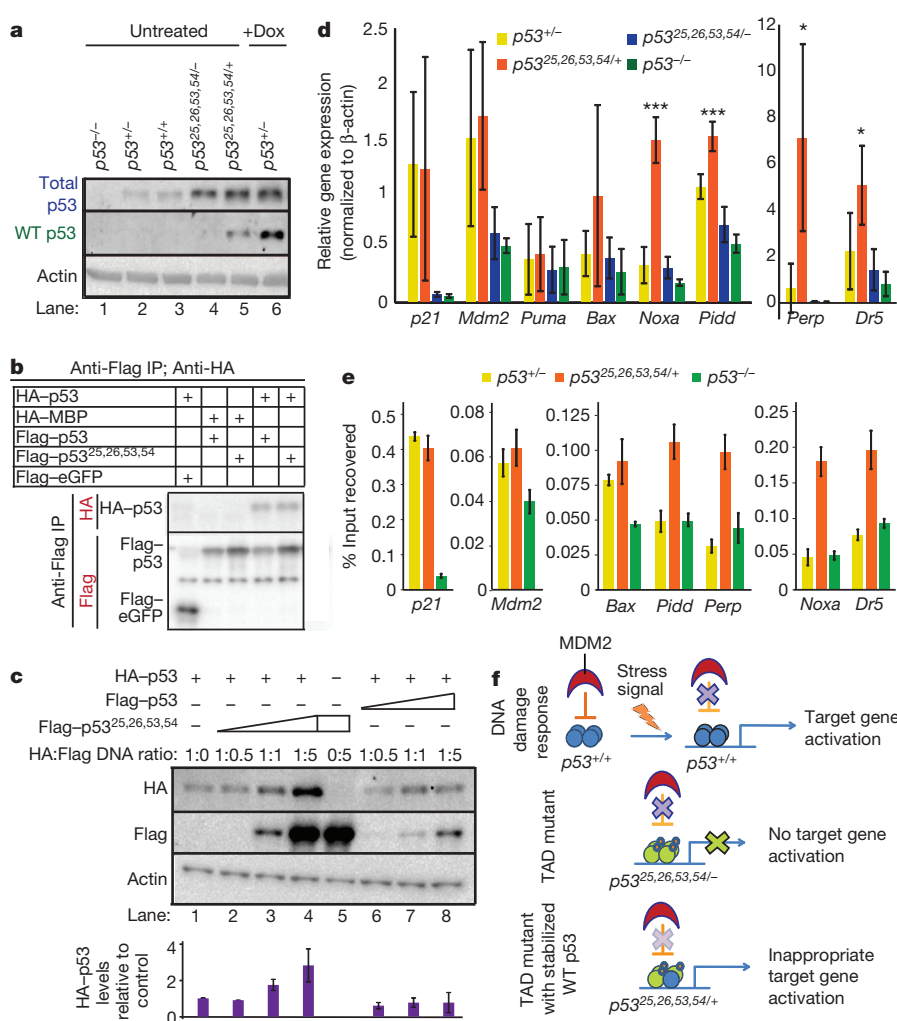


Figure 3 | $p53^{25,26,53,54}$ interacts with and increases wild-type p53 levels and activity. **a**, Immunoblot for total p53 (top row) and wild-type (WT) p53 (second row) in untreated and doxorubicin (Dox)-treated MEFs of the indicated genotypes. β -Actin was used as a loading control. **b**, Anti-Flag immunoprecipitation (IP) and immunoblot of $p53^{-/-}$ MEFs transiently overexpressing HA-p53 and Flag-p53 or Flag-p53^{25,26,53,54}. The negative controls were HA-MBP and Flag-enhanced GFP (eGFP). (See also Extended Data Fig. 7b.) **c**, Top, Immunoblot of $p53^{-/-}$ MEFs transiently overexpressing HA-p53 and increasing amounts of Flag-p53 or Flag-p53^{25,26,53,54}. Bottom, Mean \pm s.d. HA-p53 protein levels relative to lane 1 and normalized to β -actin ($n = 3$). **d**, p53 target gene expression in untreated MEFs of the indicated genotypes, presented as mean \pm s.d. normalized to β -actin gene expression ($n = 4$). *, $P < 0.05$; ***, $P < 0.005$, $p53^{25,26,53,54/+}$ versus $p53^{+/+}$ MEFs by two-tailed Student's t -test. **e**, Representative analysis from duplicate p53 ChIP analyses of p53 target genes in MEFs of the indicated genotypes relative to input DNA. **f**, Proposed model for how $p53^{25,26,53,54}$ affects p53 activity.

architecture and haematopoiesis (Extended Data Fig. 4h). Evaluation for choanal atresia and external genital defects, which are features of CHARGE syndrome, was precluded by late-gestational embryonic lethality. Importantly, $p53$ -null embryos did not display CHARGE-like phenotypes, suggesting that these phenotypes result from p53 activation^{12–14} (Extended Data Fig. 5). Collectively, the analysis of $p53^{25,26,53,54/+}$ embryonic phenotypes revealed a strong similarity to phenotypes in patients with CHARGE syndrome (Fig. 2d and Extended Data Fig. 4i), including the presence of hallmarks such as coloboma, ear malformations and heart defects. Additionally, we observed exencephaly and late-gestational lethality, neither of which is commonly reported in CHARGE syndrome. However, it remains possible that fetuses with more severe phenotypes of CHARGE syndrome die *in utero*^{15,16}.

To understand the underlying cellular basis of the $p53^{25,26,53,54/+}$ embryonic phenotypes, we examined whether p53 induced apoptosis or cell-cycle arrest in these embryos. Analysis of the retina, which is affected in coloboma, and NCCs, which are responsible for some CHARGE phenotypes, revealed more apoptosis and less proliferation in $p53^{25,26,53,54/+}$ embryos than in littermate controls (Fig. 2e, f and Extended Data Fig. 6a, c, d). Similar results were observed in other tissues affected in CHARGE syndrome, including the thymus, neuroepithelium and otic vesicles (Extended Data Fig. 6b, e, f). Thus, both increased apoptosis and reduced proliferation contribute to the $p53^{25,26,53,54/+}$ embryonic phenotypes.

We next investigated the molecular mechanisms by which $p53^{25,26,53,54}$ triggers CHARGE-associated phenotypes. Mutation of p53 residues 25 and 26 inhibits MDM2 interaction, resulting in inappropriate stabilization of the $p53^{25,26,53,54}$ protein, as seen in untreated $p53^{25,26,53,54/+}$ and $p53^{25,26,53,54/-}$ mouse embryonic fibroblasts (MEFs) compared with

untreated $p53^{+/+}$ MEFs (Fig. 3a and Extended Data Fig. 7a). Using co-transfection and co-immunoprecipitation, we found that $p53^{25,26,53,54}$ interacts with wild-type p53 (Fig. 3b and Extended Data Fig. 7b). Moreover, overexpressing increasing amounts of Flag-p53^{25,26,53,54} (lanes 2–4), but not Flag-p53 (lanes 6–8), caused haemagglutinin (HA)-p53 protein accumulation (Fig. 3c). Similarly, immunoblot analysis with a wild-type p53-specific antibody revealed that untreated $p53^{25,26,53,54/+}$ MEFs contained more wild-type p53 protein (lane 5) than did untreated $p53^{+/+}$ or $p53^{25,26,53,54/-}$ MEFs (lanes 2 and 3; Fig. 3a). To examine the effects of increased wild-type p53 abundance, we examined p53 target gene expression in $p53^{25,26,53,54/+}$ and control MEFs. $p53^{25,26,53,54}$ alone displayed no transcriptional activity, as seen in genome-wide microarray analyses³ (Extended Data Fig. 7c) and individual gene quantitative PCR with reverse transcription (qRT-PCR) assays comparing $p53^{25,26,53,54/+}$ and $p53^{-/-}$ MEFs (Fig. 3d). By contrast, $p53^{25,26,53,54}$ combined with wild-type p53 drove elevated expression of certain p53 target genes, including *Noxa* (also known as *Pmaip1*) and *Pidd*, but not *p21* (also known as *Cdkn1a*) and *Mdm2*, relative to the levels in $p53^{+/+}$ and $p53^{+/+}$ MEFs, suggesting that $p53^{25,26,53,54}$ activates wild-type p53 to induce the expression of specific p53 target genes (Fig. 3d and Extended Data Fig. 7d). Similarly, overexpression of $p53^{25,26,53,54}$, but not wild-type p53, in $p53^{+/+}$ MEFs significantly enhanced the expression of certain p53 target genes (Extended Data Fig. 7e). Quantitative chromatin immunoprecipitation (ChIP) analyses revealed that this selective target gene activation resulted from greater binding of p53 to the response elements of particular target genes in $p53^{25,26,53,54/+}$ MEFs than in $p53^{+/+}$ MEFs (Fig. 3e). This pattern probably reflects different affinities of p53-binding sites for p53 such that p53 can only bind to, and activate

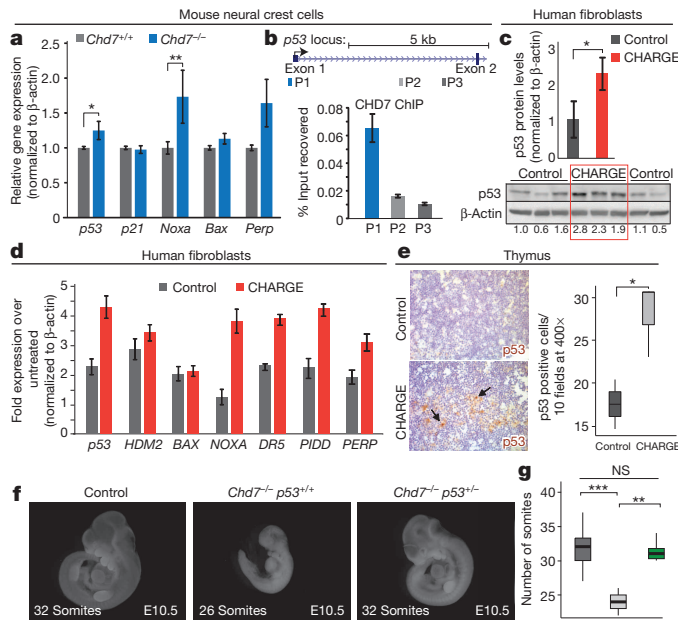


Figure 4 | p53 is activated upon CHD7 deficiency and contributes to *Chd7*-null phenotypes. **a**, *p53* target gene expression in *Chd7*^{+/+} and *Chd7*^{-/-} (*whi/whi*) NCCs, presented as mean \pm s.e.m. normalized to β -actin gene expression ($n = 5$). *, $P < 0.05$; **, $P < 0.01$, one-tailed Mann-Whitney test. **b**, Top, 5' End of *p53* locus. Bottom, Representative analysis from triplicate ChIP analyses of CHD7 binding to the *p53* promoter (P1) and to negative control open-chromatin regions (P2 and P3), relative to input \pm s.d. of technical triplicates. kb, kilobases. **c**, Bottom, *p53* immunoblot of human fibroblasts derived from patients with CHARGE syndrome and unaffected individuals (control). The numbers indicate the *p53* levels after normalization to β -actin. Top, Data are presented as mean \pm s.d. of *p53* levels. *, $P = 0.015$, one-tailed Student's *t*-test. **d**, *p53* target gene expression in low-serum treated human fibroblasts derived from patients with CHARGE syndrome and unaffected individuals analysed in duplicate. A representative cell line pair is shown, and data are presented as mean \pm s.d. of technical triplicates. **e**, Left, *p53* immunohistochemistry on thymi from fetuses with CHARGE syndrome and unaffected fetuses. Arrows denote *p53*-positive cells. Right, The number of *p53*-positive cells per ten fields at 400 \times magnification, from one section; $n = 3$. *, $P = 0.037$, one-tailed Student's *t*-test. **f**, Comparison of *Chd7*^{-/-} *p53*^{+/+} embryos with control and *Chd7*^{-/-} *p53*^{+/+} embryos reveals partial rescue of E10.5 *Chd7*^{-/-} *p53*^{+/+} embryos (*Chd7*^{-/-} *p53*^{+/+}, $n = 3$; *Chd7*^{-/-} *p53*^{+/+}, $n = 6$). Original magnification, $\times 20$. **g**, Embryo somite number for each genotype. **, $P = 0.0024$; ***, $P = 0.00032$; NS, not significant; two-tailed Student's *t*-test.

transcription from, lower affinity *p53*-binding sites (such as those in *Noxa* and *Pidd*) when sufficiently stabilized, as is the case in the presence of both *p53*^{25,26,53,54} and wild-type *p53* (ref. 17). Collectively, these findings suggest that in *p53*^{25,26,53,54/+} embryos, *p53*^{25,26,53,54} interacts with and stabilizes wild-type *p53*, probably by compromising the interaction of *p53* with MDM2, causing *p53* to inappropriately induce target gene expression, cell-cycle arrest and/or apoptosis, and CHARGE-like phenotypes (Fig. 3f).

These observations suggested that *p53* may respond to CHD7 deficiency and play a role in CHARGE syndrome. Analysis of *Chd7* expression levels in *p53*^{+/+} and *p53*^{25,26,53,54/+} MEFs revealed no significant difference, suggesting that altered *Chd7* levels do not underlie the CHARGE-like phenotypes in *p53*^{25,26,53,54/+} embryos (Extended Data Fig. 8a). We next assessed whether *p53* responds to changes in CHD7 status^{15,16,18,19}. Indeed, *p53* and certain *p53* target genes were induced in *Chd7*-null NCCs (Fig. 4a and Extended Data Fig. 8b). Interestingly, ChIP analyses showed that CHD7 binds to the *p53* promoter in NCCs, suggesting that CHD7 negatively regulates *p53* expression, providing a

mechanism by which CHD7 loss might contribute to a *p53* response (Fig. 4b), although not excluding the possibility that CHD7 deficiency could also activate *p53* through other mechanisms. To directly assess *p53* activation in CHARGE syndrome, we analysed *CHD7*-mutation-positive fibroblasts from patients with CHARGE syndrome and found higher basal *p53* protein levels than in controls, as well as *p53* target gene induction, following mild stress (Fig. 4c, d). Similarly, analysis of thymi showed more *p53*-positive thymocytes in *CHD7*-mutation-positive fetuses with CHARGE syndrome than in unaffected fetuses (Fig. 4e). Thus, *p53* is activated in *Chd7*-null NCCs and in fibroblasts and tissue from patients with CHARGE syndrome. To establish the role of *p53* downstream of *Chd7* loss, we tested whether *Chd7*-null phenotypes are rescued by *p53* heterozygosity. We found that the characteristic *Chd7*-null phenotypes of severe developmental delay and generalized hypoplasia at E10.5 are significantly rescued on a *p53*^{+/+} background. Specifically, heart development and somite number, an indicator of developmental stage, are rescued, while limb, forebrain and facial morphogenesis are partially rescued (Fig. 4f, g and Extended Data Fig. 8c, d). Thus, *p53* heterozygosity rescues phenotypes caused by CHD7 inactivation, although incompletely, which is consistent with the existence of both *p53*-dependent and *p53*-independent responses downstream of CHD7 loss. Collectively, these findings demonstrate that CHD7 deficiency triggers *p53* activation and *p53*-dependent phenotypes.

Our *p53*^{25,26,53,54/+} mouse strain provides a new model with which to study features that are relevant to CHARGE syndrome. Mouse models of CHARGE syndrome have been generated previously by using *N*-ethyl-*N*-nitrosourea (ENU) mutagenesis or by targeting the *Chd7* locus^{20,21}. While *Chd7*^{-/-} mutants display embryonic lethality at approximately E10.5, *Chd7*^{+/+} mutants are viable, exhibiting defects of the inner ear, heart, external genitalia and choanae/palate^{20,21}. In the inner ear, *Chd7*^{+/+} mice display truncated lateral SCCs and variably truncated posterior SCCs^{20,22}, whereas *p53*^{25,26,53,54/+} embryos primarily exhibit posterior SCC defects or defects in all three SCCs, similarly to patients with CHARGE syndrome, in which all three SCCs are often involved²³. Interestingly, *Chd7*^{+/+} mice have not been reported to display the CHARGE hallmarks of coloboma or cardiac outflow tract defects. Thus, the presence of extensive ear defects, heart defects and coloboma in *p53*^{25,26,53,54/+} embryos highlights the utility of our model for recapitulating a broader form of CHARGE syndrome, potentially representing a phenotype that is intermediate in severity between *Chd7*^{+/+} and *Chd7*^{-/-} mutants.

Interestingly, the set of phenotypes of *p53*^{25,26,53,54/+} embryos is one set in a spectrum of mouse models whereby varying levels of *p53* activity trigger different phenotypes. Unlike *Mdm2* deficiency, which results in stabilized, fully active wild-type *p53*, the combination of wild-type *p53* and *p53*^{25,26,53,54} in *p53*^{25,26,53,54/+} embryos results in modest *p53* activation, causing lethality and phenotypes less severe than *Mdm2* loss²⁴. The *p53*^{25,26,53,54/+} embryonic phenotypes are also in contrast to those seen in other hyperactive *p53* mouse models, including one model in which the expression of a carboxy-terminal *p53* fragment causes premature ageing and another model in which reduced MDM2 expression triggers lymphopaenia and reduced body weight in adults^{25–27}. In these models, the phenotypes manifested without enhanced basal *p53* activity or increased *p53* stabilization, respectively, potentially explaining the lack of embryonic lethality and the milder phenotypes than observed for *p53*^{25,26,53,54/+} embryos.

Mechanistically, our observation that *p53* activation is sufficient to cause CHARGE-like phenotypes in mice suggested that activated *p53* could similarly promote the characteristic defects present in human CHARGE syndrome. Indeed, *p53* expression and activity are increased in the presence of CHD7 deficiency, and phenotypes triggered by CHD7 loss are partially rescued by *p53* heterozygosity. This finding is in contrast to a study in which *p53* morpholinos failed to rescue *Chd7* deficiency in zebrafish, perhaps reflecting species-specific differences²⁸. *p53* may also function independently of CHD7 to induce the features of CHARGE syndrome, a possibility that is relevant in the ~10–30% of CHARGE cases in which CHD7 is not mutated. *p53* could become

activated in response to other genetic alterations found in CHARGE syndrome, or components of the p53 pathway could themselves be mutated in some CHARGE cases. Our sequencing analysis of the p53 coding region in 25 *CHD7*-mutation-negative patients with CHARGE failed to reveal any mutations, potentially because modest p53 activation may be difficult to achieve through point mutation, suggesting that mutations in p53 *cis*-regulatory regions or pathway components may be more likely. Significantly, p53 may act as a common node for developmental defects not only in CHARGE syndrome but also in other developmental syndromes, such as 22q11.2 deletion syndrome²⁹. Indeed, p53 activation by ribosome dysfunction induces NCC deficiency and craniofacial defects³⁰. Future studies will reveal precisely how p53 contributes to CHARGE syndrome and potentially the broader spectrum of human craniofacial and cardiac developmental syndromes.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 July 2013; accepted 16 June 2014.

Published online 3 August 2014.

- Davenport, S. L. H., Hefner, M. A. & Mitchell, J. A. The spectrum of clinical features in CHARGE syndrome. *Clin. Genet.* **29**, 298–310 (1986).
- Jongmans, M. C. J. *et al.* CHARGE syndrome: the phenotypic spectrum of mutations in the *CHD7* gene. *J. Med. Genet.* **43**, 306–314 (2006).
- Brady, C. A. *et al.* Distinct p53 transcriptional programs dictate acute DNA-damage responses and tumor suppression. *Cell* **145**, 571–583 (2011).
- Marine, J. C. *et al.* Keeping p53 in check: essential and synergistic functions of Mdm2 and Mdm4. *Cell Death Differ.* **13**, 927–934 (2006).
- Johnson, T. M., Hammond, E. M., Giaccia, A. & Attardi, L. D. The p53Qs transactivation-deficient mutant shows stress-specific apoptotic activity and induces embryonic lethality. *Nature Genet.* **37**, 145–152 (2005).
- Brock, K. E., Mathiason, M. A., Rooney, B. L. & Williams, M. S. Quantitative analysis of limb anomalies in CHARGE syndrome: correlation with diagnosis and characteristic CHARGE anomalies. *Am. J. Med. Genet. A* **123A**, 111–121 (2003).
- Zentner, G. E., Layman, W. S., Martin, D. M. & Scacheri, P. C. Molecular and phenotypic aspects of *CHD7* mutation in CHARGE syndrome. *Am. J. Med. Genet. A* **152A**, 674–686 (2010).
- Lalani, S. R. *et al.* Spectrum of *CHD7* mutations in 110 individuals with CHARGE syndrome and genotype–phenotype correlation. *Am. J. Hum. Genet.* **78**, 303–314 (2006).
- Ragan, D. C., Casale, A. J., Rink, R. C., Cain, M. P. & Weaver, D. D. Genitourinary anomalies in the CHARGE association. *J. Urol.* **161**, 622–625 (1999).
- Inoue, H. *et al.* Successful cord blood transplantation for a CHARGE syndrome with *CHD7* mutation showing DiGeorge sequence including hypoparathyroidism. *Eur. J. Pediatr.* **169**, 839–844 (2010).
- Corsten-Janssen, N. *et al.* The cardiac phenotype in patients with a *CHD7* mutation. *Circ. Cardiovasc. Genet.* **6**, 248–254 (2013).
- Rinon, A. *et al.* p53 coordinates cranial neural crest cell growth and epithelial–mesenchymal transition/delamination processes. *Development* **138**, 1827–1838 (2011).
- Armstrong, J. F., Kaufman, M. H., Harrison, D. J. & Clarke, A. R. High-frequency developmental abnormalities in p53-deficient mice. *Curr. Biol.* **5**, 931–936 (1995).
- Lengner, C. J. *et al.* Osteoblast differentiation and skeletal development are regulated by Mdm2–p53 signaling. *J. Cell Biol.* **172**, 909–921 (2006).
- Sanlaville, D. *et al.* Phenotypic spectrum of CHARGE syndrome in fetuses with *CHD7* truncating mutations correlates with expression during human development. *J. Med. Genet.* **43**, 211–217 (2006).
- Legendre, M. *et al.* Antenatal spectrum of CHARGE syndrome in 40 fetuses with *CHD7* mutations. *J. Med. Genet.* **49**, 698–707 (2012).
- Veprintsev, D. B. & Fersht, A. R. Algorithm for prediction of tumour suppressor p53 affinity for binding sites in DNA. *Nucleic Acids Res.* **36**, 1589–1598 (2008).
- Bajpai, R. *et al.* *CHD7* cooperates with PBAF to control multipotent neural crest formation. *Nature* **463**, 958–962 (2010).
- Schmetz, M. P. *et al.* *CHD7* targets active gene enhancer elements to modulate ES cell-specific gene expression. *PLoS Genet.* **6**, e1001023 (2010).
- Bosman, E. A. *et al.* Multiple mutations in mouse *Chd7* provide models for CHARGE syndrome. *Hum. Mol. Genet.* **14**, 3463–3476 (2005).
- Hurd, E. *et al.* Loss of *Chd7* function in gene-trapped reporter mice is embryonic lethal and associated with severe defects in multiple developing tissues. *Mamm. Genome* **18**, 94–104 (2007).
- Adams, M. E. *et al.* Defects in vestibular sensory epithelia and innervation in mice with loss of *Chd7* function: implications for human CHARGE syndrome. *J. Comp. Neurol.* **504**, 519–532 (2007).
- Morimoto, A. K. *et al.* Absent semicircular canals in CHARGE syndrome: radiologic spectrum of findings. *AJNR Am. J. Neuroradiol.* **27**, 1663–1671 (2006).
- de Oca Luna, R. M., Wagner, D. S. & Lozano, G. Rescue of early embryonic lethality in *mdm2*-deficient mice by deletion of *p53*. *Nature* **378**, 203–206 (1995).
- Mendrysa, S. M. *et al.* *mdm2* is critical for inhibition of p53 during lymphopoiesis and the response to ionizing irradiation. *Mol. Cell. Biol.* **23**, 462–472 (2003).
- Mendrysa, S. M. *et al.* Tumor suppression and normal aging in mice with constitutively high p53 activity. *Genes Dev.* **20**, 16–21 (2006).
- Tyner, S. D. *et al.* p53 mutant mice that display early ageing-associated phenotypes. *Nature* **415**, 45–53 (2002).
- Balow, S. A. *et al.* Knockdown of *fbx10/kdm2bb* rescues *chd7* morphant phenotype in a zebrafish model of CHARGE syndrome. *Dev. Biol.* **382**, 57–69 (2013).
- Corsten-Janssen, N. *et al.* More clinical overlap between 22q11.2 deletion syndrome and CHARGE syndrome than often anticipated. *Mol. Syndromol.* **4**, 235–245 (2013).
- Jones, N. C. *et al.* Prevention of the neurocristopathy Treacher Collins syndrome through inhibition of p53 function. *Nature Med.* **14**, 125–133 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Spano-Mello, K. T. Biegging, N. Raj and M. Monje-Deisseroth for reading the manuscript and S. E. Artandi and T. Williams for discussion. We thank H. Chou for immunohistochemistry assistance; E. L. Van Nostrand, P. Lavori, and A. McMillian for statistical analysis; K. Weinberg and D. Min for thymus analysis assistance; M. Shkrelil for kidney analysis assistance; B. Liu and J. A. Helms for craniofacial analysis assistance; and M. Bowen for *Chd7* mouse experiment assistance. We thank S. E. Artandi for plasmids; S. E. Artandi and P. Khavari for control human fibroblast cell lines; D. Lane and B. Vojtesek for wild-type p53-specific antibody (pAB242); P. Scacheri for wild-type and *Chd7*-null mouse embryonic stem cells; and T. Denecker and G. Goudefroye for *TP53* sequencing in patients. This work was supported by funding from the NSF and NCI (grant number 1F31CA167917-01) to J.L.V.N.; from the NIH (R01 GM095555) to J.W.; from the American Heart Association (12EIA8960018), March of Dimes Foundation (#6-FY11-260) and NIH (R01 HL118087 and R01 HL121197) to C.-P.C.; from the NIH (R01 DC009410) to D.M.M.; and from the ACS, LLS and NIH (R01 CA140875) to L.D.A.

Author Contributions J.L.V.N. designed and carried out experiments, interpreted data and wrote the manuscript. C.A.B. generated the p53^{25,26,53,54} mice, designed and carried out experiments, and interpreted data. H.J. performed p53 ChIP analyses. M.M.K. and T.M.J. performed certain mouse analyses. D.R.F. and J.W. performed NCC differentiation and *CHD7* ChIP analyses. C.-Y.L., C.-J.L. and C.-P.C. assisted with heart analyses. D.L.S. and D.M.M. performed inner ear analyses, interpreted data and provided human fibroblasts. H.V. assisted with histological analyses. J.A.B. generated CHARGE and control human fibroblast lines. T.A.-B. performed *TP53* sequencing analysis in patients and supplied CHARGE thymus samples. L.D.A. designed experiments, interpreted data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.D.A. (attardi@stanford.edu).

METHODS

Mouse breeding and analysis. Conditional *p53* mutant mice were described previously^{3,5}. *p53*^{LSL-mut/+} males were crossed to *CMV-Cre* females, and timed pregnancies were conducted³¹. Genotyping analysis was performed using yolk sac DNA. Males were identified using the PCR for the Y-chromosome-specific *Zfy* gene. As reported³¹, we observed mosaic Cre activity in some embryos, reflected by incomplete or no *lox-Stop-lox* deletion in PCR analysis of yolk sacs and in *p53* immunohistochemical analysis in embryos. Embryos that showed little to no recombination of the *lox-Stop-lox* allele were kept separate for all analyses. In the text and figure labels, the Cre nomenclature for both control and *p53*^{25,26,53,54/+} embryos is excluded for simplicity. The genotypes of *p53*^{25,26,53,54/+} mice carrying a *CMV-Cre* transgene lack the LSL designation because the *lox-Stop-lox* element has been deleted from the genome. Controls in analyses comprise littermate embryos both with and without the *CMV-Cre* transgene (Extended Data Fig. 3). Mice were maintained on a mixed 129/Sv; C57BL/6J background. *Chd7*-deficient gene-trapped mice were described previously²¹. Controls for embryo somite number rescue analysis comprised *Chd7*^{+/+}*p53*^{+/+}, *Chd7*^{+/+}*p53*^{+/-}, *Chd7*^{+/-}*p53*^{+/+} and *Chd7*^{+/-}*p53*^{+/-} embryos. Sample sizes were estimated, based on previous embryogenesis work, to be able to reach significant conclusions. For example, using a chi-squared test, we can estimate that 24 embryos are required to obtain a significance level of 5% if all homozygous embryos were non-viable and from a heterozygous mating where homozygotes should represent 25% of the total based on Mendelian ratio. All animal work was done in accordance with the Stanford University APLAC.

Embryo tissue analysis. Embryos were examined under a dissecting microscope for the presence of a heartbeat and for other abnormalities, in a blinded fashion, before genotyping, and images of either fixed or live embryos were acquired using the dissecting microscope. Whether mice were alive was determined by the presence of a heartbeat. All analyses were performed on embryos of specified ages as determined by the expected stage of tissue and organ development in control embryos. Within age groups, embryos were randomly assigned for the analysis of individual tissues. Heart, inner ear and craniofacial analyses were performed with the assistance of collaborators who were blinded to the genotypes. Histological analysis was performed on haematoxylin and eosin stained paraffin-embedded sagittal, coronal or transverse embryo sections using standard protocols. Consecutive 7-µm sections were evaluated for heart and craniofacial analyses. Whole-mount images for craniofacial analysis were created by staining embryos in 70% ethanol with ethidium bromide and imaging with ultraviolet radiation. Whole-mount cleaved caspase 3 (CC3) staining was performed as described previously³² with anti-CC3 antibody (Cell Signaling Technology, 9664), and embryos were developed with diaminobenzidine (DAB) (Vector Labs). Immunohistochemistry and immunofluorescence were performed as described with anti-CC3 antibody, anti-5-bromodeoxyuridine (BrdU) antibody (BD Bioscience, 347580), anti-*p53* antibody (Vector Labs, CM5), anti-PAX3 antibody (Iowa Developmental Studies Hybridoma Bank) and human anti-*p53* antibody (Santa Cruz Biotechnology, Ab1801) on paraformaldehyde-fixed, paraffin-embedded tissue, and slides were developed with either DAB or imaged by immunofluorescence microscopy. For BrdU staining, the embryos were pulsed before dissection for 20 min with 0.1 mg BrdU per g body weight. CC3- and BrdU-positive cells were quantified and normalized to the area, number of sections or total cell number. Bone and cartilage staining were performed as described previously³³, using Alizarin red and Alcian blue to stain bone and cartilage, respectively. Bone lengths were quantified from measurements on photomicrographs acquired at 6.3× magnification. The areas of the kidney and thymus were calculated from measurements on photomicrographs acquired at 200× magnification. Inner ear analysis using the paint-fill assay was performed as described previously³⁴; paint-filled ears were imaged in brightfield using a Leica DMRB stereoscope. All individuals with affected ears that could be scored for both ears exhibited a similar severity of the defect in both ears. Additionally, while different types of analysis precluded investigation of all phenotypes in particular embryos, defects commonly occurred in combination: for example, coloboma and inner ear defects occurred together in 57% of *p53*^{25,26,53,54/+} embryos examined.

qRT-PCR. For qRT-PCR, MEFs were cultured at subconfluency; RNA was isolated by TRIzol extraction and reverse transcribed using M-MLV Reverse Transcriptase (Invitrogen) and random primers; and PCR was performed in triplicate using the SYBR Green system (QIAGEN and Bio-Rad) and mouse-specific primers for each gene (Supplementary Methods Table 1a) in a 7900HT Fast Real-Time PCR machine (Applied Biosystems). The results were computed relative to a standard curve made with cDNA pooled from all samples³. Human fibroblasts were cultured at subconfluency in DMEM with 20% serum or 0.1% serum for 24 h before RNA isolation by TRIzol extraction, and qRT-PCR was performed as above using the SYBR Green system and human-specific primers for each gene (Supplementary Methods Table 1b).

Cell culture, western blot analysis and immunoprecipitation. MEFs were derived from E13.5 embryos. For stable overexpressing MEF lines, wild-type and *p53*-null

MEFs were transduced with pWZL-based retroviruses expressing Flag-*p53*, Flag-*p53*^{25,26,53,54} or empty vector as described previously⁵. Mutant *p53* constructs were made via site-directed mutagenesis to bear Leu25Gln, Trp26Ser, Phe53Gln and Phe54Ser mutations, which were found to incapacitate *p53*'s transactivation capability. In summary, a *p53* cDNA, generated from MEF RNA, was amplified and engineered with *AscI* and *PacI* restriction enzyme sites using primers against the second amino acid through the stop codon (5'-TTAGGCGCGCCACTGCCATGGAGGAGTC-3' and 5'-GGCGTAAATTAATCAGTCTGAGTCAGGCCCCAC-3'). The amplified DNA was subcloned into a pBluescript construct with the multiple cloning site replaced with *AscI* and *PacI* sites. The *p53* cDNA underwent site-directed mutagenesis (TTATGG to CAATCG and TTTT to CAGTCT) to generate the four mutations. The mutated *p53* cDNA was confirmed by sequencing and subcloned into the appropriate expression vectors (pWZL or pcDNA) that contained three copies of either the HA or Flag tag at the amino terminus, as well as *AscI* and *PacI* sites instead of the multiple cloning sites. For western blot analysis, cells were left untreated, treated with 0.2 µg ml⁻¹ doxorubicin (Dox) for 8 h or transfected using X-tremeGENE HP (Roche) for 24 h with pcDNA constructs (Flag-*p53*, Flag-*p53*^{25,26,53,54}, HA-*p53* or empty vector). Cells were collected and lysed using RIPA buffer. Western blots were probed with anti-*p53* (Vector Labs, CM5, 1:500), anti-wild-type *p53* (provided by D. Lane and B. Vojtesek, pAB242)³⁵, anti-Flag (Sigma, F3165, 1:1,000), anti-HA (Sigma, H6908, 1:1,000) and anti-actin (Sigma, A2228, 1:30,000) antibodies. Immunoprecipitation was performed on MEFs transfected with pcDNA constructs (Flag-*p53*, Flag-*p53*^{25,26,53,54}, HA-*p53*, Flag-eGFP or HA-MBP) using M2 anti-Flag agarose beads (Sigma, A2220). Dox-treated *p53*^{+/+} (Fig. 3a, lane 6) and *p53*^{25,26,53,54/+} (Fig. 3a, lane 4) MEFs controlled for wild-type *p53*-specific antibody activity and specificity, respectively. Cells were lysed using RIPA buffer; lysates were added to anti-Flag beads to allow the binding of Flag protein; and bound protein was retrieved by the addition of sample buffer and the centrifugation of beads. Ten per cent input was used for western blot analysis. Human fibroblasts were grown in DMEM with 20% serum, and were lysed using RIPA buffer or were incubated in 0.1% serum for 24 h before collection for RNA analysis. Western blots were probed with anti-*p53* antibody (Santa Cruz Biotechnology, DO-1). The differentiation of mouse *Chd7*^{+/+} and *Chd7*^{-/-} (*whi/whi*) embryonic stem cells into NCCs was performed using an adaptation of a previously described protocol^{18,19} that has been further characterized³⁶. Importantly, loss of CHD7 function does not affect the induction of NCCs, allowing for generation of the cells for this analysis¹⁸. Validation of NCC phenotype was performed by assessing the gene expression of the NCC markers *Wnt1*, *Snai1* and *Pax3* (Supplementary Methods Table 1a).

ChIP. Analysis of *p53* binding was performed on MEFs of different genotypes. For ChIP, 30–40 × 10⁷ cells were collected, and immunoprecipitation was conducted using anti-*p53* antibody (Vector Labs, CM5), as previously described³⁷. In brief, the cells were crosslinked following trypsinization with 1% formaldehyde and quenched with glycine. Chromatin was sheared using a Bioruptor sonicator (Diagenode), then immunoprecipitated with anti-*p53*-Dynabeads overnight. Immunoprecipitated DNA was washed, reverse-crosslinked and isolated using a PCR clean-up kit (QIAGEN). Immunoprecipitated DNA was analysed using qRT-PCR with binding-site-specific primers (Supplementary Methods Table 2) and normalized to input. Analysis of *p53* binding-site affinities was based on previously reported ChIP-sequencing data and a *p53* binding-site algorithm^{17,37}. Analysis of CHD7 binding was performed using anti-CHD7 antibody on NCC-like cells differentiated from wild-type or *Chd7*-null (*whi/whi*) embryonic stem cells using the same protocol as above¹⁸.

Human samples. All human work was conducted under human subject protocols approved by the Stanford Institutional Review Board (IRB), the University of Michigan UM-IRBMED and the Ethical Committee of d'Ile de France II (N° 2009-164). Informed consent was obtained from participants in the study. Human fibroblasts were obtained by skin biopsy under local anaesthesia from CHARGE patients with *CHD7* mutations and unaffected individuals. Thymi were obtained from pregnancies terminated for severe malformations, in accordance with French law, after genetic counselling, between 24 and 37 weeks of gestation. Detailed clinicopathological examination allowed the diagnosis of CHARGE syndrome, which was confirmed by *CHD7* molecular analysis that identified a *de novo* nonsense mutation in each sample. Age-matched control thymi were obtained from fetuses with isolated brain malformations.

TP53 (*p53*) exome sequence was analysed in a group of 25 *CHD7*-mutation-negative patients with features of CHARGE syndrome, who were selected in the cohort based on the availability of parental DNA for segregation analysis, in case a variant was found. All *TP53* coding exons were analysed by direct sequencing of PCR products comprising the ten coding exons and the adjacent intronic junctions of *TP53* isoform 2 (exons 2 to 11, reference sequence NM_001126112.2) and the alternatively spliced exon 10 of isoforms 3 and 4 (reference sequences NM_001126114.2 and NM_001126113.2). This analysis failed to reveal any mutations within the coding

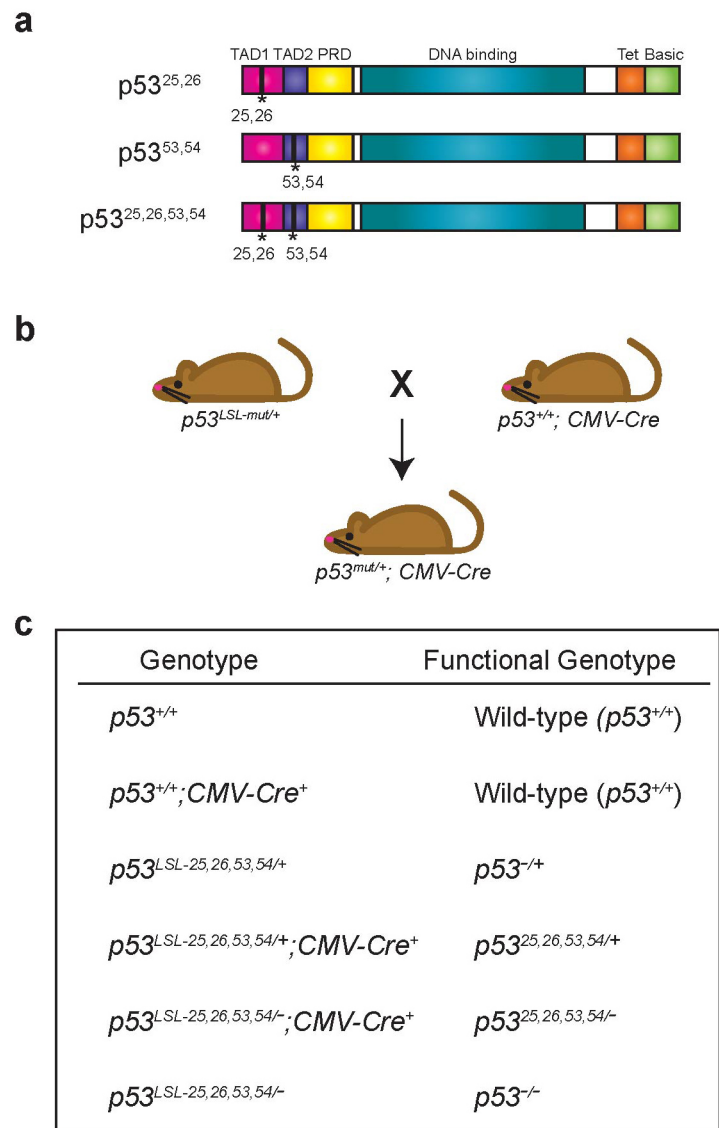
region (that is, no sequence changes apart from already known single nucleotide polymorphisms).

Statistical analysis. The statistical tests used were Student's *t*-test (equal variance), Welch's *t*-test (unequal variance), the Mann–Whitney test (non-parametric test) and the binomial distribution test. Owing to the presence of embryos with little or no recombination of the *lox-Stop-lox* allele, binomial distribution analysis was used to exclude these embryos and to determine the statistical significance for the viability of $p53^{25,26,53,54/+}$ embryos at each embryonic time point. Additional statistical analyses, with the assistance of the Stanford Biostatistics Core, were performed to assess viability over multiple embryonic time points of *CMV-Cre*; $p53^{25,26,53,54/+}$ embryos that had recombined versus those that had not recombined, as scored by PCR analysis (Extended Data Table 2).

The number of dead $p53^{25,26,53,54/+}$ embryos, as a proportion of the total (dead plus alive), was also analysed using logistic regression, with a factor for genotype and a factor for gestational age. Because some of the totals were small, we assessed significance using a permutation test, comparing the observed phenotype (the ratio of dead embryos) to a set of 10,000 random permutations of the phenotype (keeping the number of dead embryos and the total numbers unchanged at each time point and re-fitting the logistic regression). Note that this analysis assumes the effect of clustering within dams is negligible. The *P* value for the permutation test was 0.0234: 234 out of 10,000 random permutations showed a greater frequency of death in absolute value than the frequency of death observed for $p53^{25,26,53,54/+}$ embryos; the *P* value from logistic regression software was *P* = 0.0003. A model

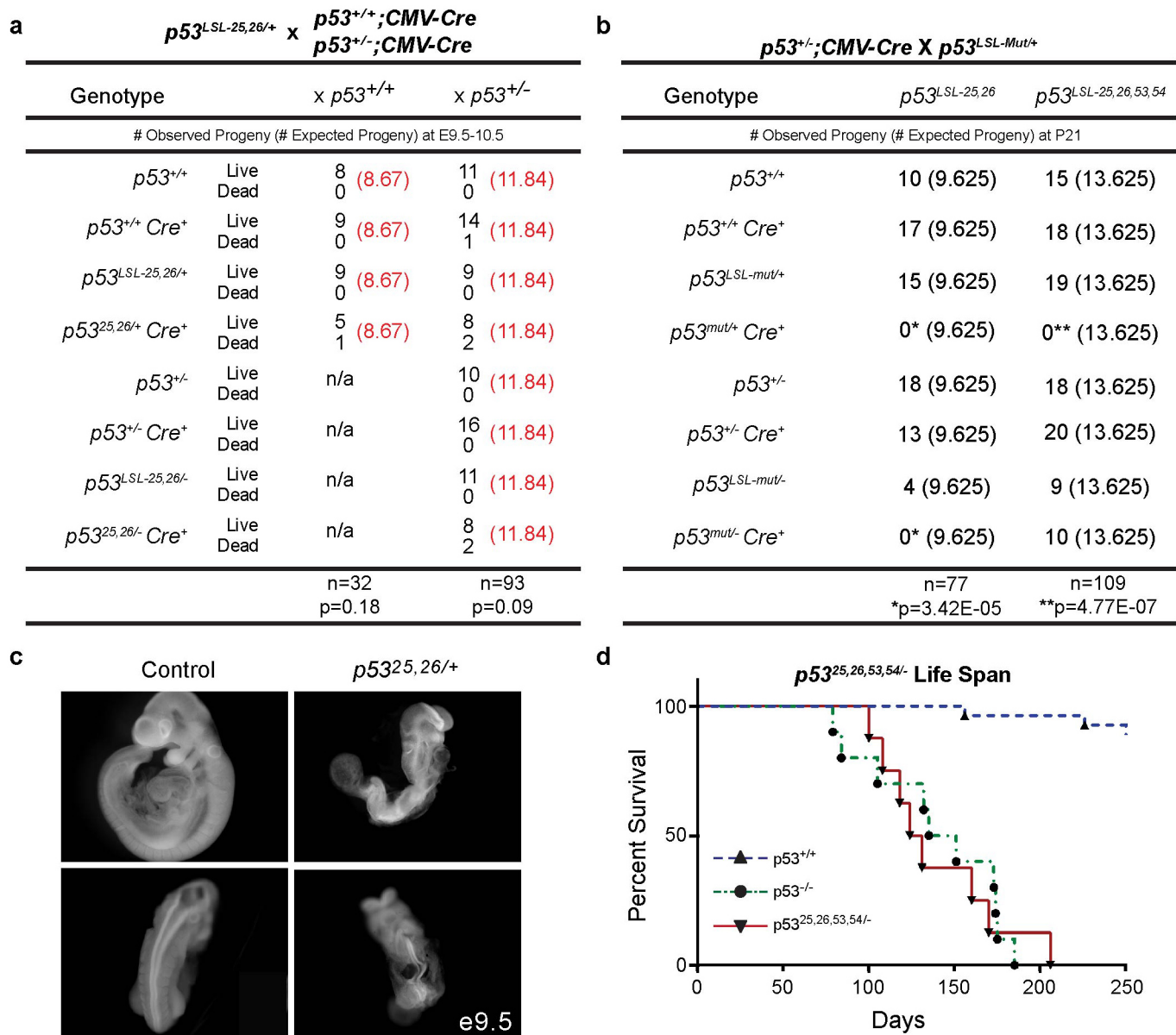
allowing a continuous effect of gestational age was also fitted to explore whether the genotype affected the rate of change in the hazard of death: the slopes over time for $p53^{25,26,53,54/+}$ and control embryos were significantly different (*P* = 0.0245), despite the confidence intervals at each individual time point being wide and overlapping.

31. Schwenk, F., Baron, U. & Rajewsky, K. A *cre*-transgenic mouse strain for the ubiquitous deletion of *loxP*-flanked gene segments including deletion in germ cells. *Nucleic Acids Res.* **23**, 5080–5081 (1995).
32. Sato, T. & Bartunkova, S. Analysis of embryonic vascular morphogenesis. *Methods Mol. Biol.* **137**, 223–233 (2000).
33. Ovchinnikov, D. Alcian blue/alizarin red staining of cartilage and bone in mouse. *Cold Spring Harb. Protoc.* <http://dx.doi.org/10.1101/pdb.prot5170> (March 2009).
34. Hurd, E. A., Poucher, H. K., Cheng, K., Raphael, Y. & Martin, D. M. The ATP-dependent chromatin remodeling enzyme CHD7 regulates pro-neural gene expression and neurogenesis in the inner ear. *Development* **137**, 3139–3150 (2010).
35. Lane, D. *et al.* Epitope analysis of the murine p53 tumour suppressor protein. *Oncogene* **12**, 2461–2466 (1996).
36. Rada-Iglesias, A. *et al.* Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. *Cell Stem Cell* **11**, 633–648 (2012).
37. Kenzelmann Broz, D. *et al.* Global genomic profiling reveals an extensive p53-regulated autophagy program contributing to key p53 responses. *Genes Dev.* **27**, 1016–1031 (2013).



Extended Data Figure 1 | Model for examining p53-associated developmental phenotypes. **a**, Schematic of $p53^{25,26}$, $p53^{53,54}$ and $p53^{25,26,53,54}$ mutant p53 proteins. Basic, basic-residue-rich domain; PRD, proline-rich domain; TAD, transcriptional activation domain 1 or 2; Tet, tetramerization domain. **b**, $p53^{LSL-mut/+}$ mice (where *mut* can denote any of the p53 TAD mutants) were crossed to $p53^{+/+}; CMV-Cre$ mice, which express Cre in the germline, to assess the viability and developmental phenotypes of the p53-mutant-expressing progeny. **c**, Table summarizing the actual genotypes and

ultimate functional genotypes of progeny from crosses of $p53^{LSL-25,26,53,54/+}$ and $p53^{+/+}; CMV-Cre$ mice, as used throughout the manuscript. While $p53^{LSL-25,26,53,54/+}; CMV-Cre$ is the actual initial genotype, when Cre acts to delete the *lox-Stop-lox* (LSL) cassette, the genotype is written as $p53^{25,26,53,54/+}$ to reflect this recombination. In the text and figure labels, the Cre nomenclature for both control and $p53^{25,26,53,54/+}$ embryos is excluded for simplicity. Controls for analyses comprised embryos both with and without the *CMV-Cre* transgene, as summarized in Extended Data Fig. 3.



Extended Data Figure 2 | $p53^{25,26,53,54/-}$ mice, but not $p53^{25,26/+}$ or $p53^{25,26/-}$ mice, are viable. a, Crosses of $p53^{LSL-25,26/+}$ with $p53^{+/+};CMV-Cre$ or $p53^{+/-};CMV-Cre$ mice revealed a decrease in viable pups expressing $p53^{25,26}$ at E9.5–E10.5. The observed numbers of live and dead pups compared with the expected numbers of live pups are indicated: [Observed (Expected)]. The genotypes of the $p53^{25,26/+}$ and $p53^{25,26/-}$ mice carrying a $CMV-Cre$ transgene lack the LSL designation because the $lox-Stop-lox$ element has been deleted from the genome. Significance was assessed by binomial distribution statistical tests on live pups: $P = 0.18$ for the $p53^{+/+}$ cross and 0.09 for the $p53^{+/-}$ cross. **b,** Crosses of $p53^{LSL-25,26/+}$ or $p53^{LSL-25,26,53,54/+}$ with $p53^{+/-};CMV-Cre$ mice revealed that $p53^{25,26,53,54/-}$ mice, but not $p53^{25,26/-}$ mice, are viable, as assessed at postnatal day 21 (P21). *mut* denotes either mutant allele. The observed numbers of pups are indicated and compared with the expected numbers of pups: [Observed (Expected)]. The genotypes of

$p53^{mut/+}$ and $p53^{mut/-}$ mice carrying a $CMV-Cre$ transgene lack the LSL designation because the $lox-Stop-lox$ element has been deleted from the genome. The lack of pups is significant at P21 as assessed by binomial distribution statistical tests on live pups compared with expected: $p53^{25,26/+}$ and $p53^{25,26/-}$, * $P = 3.42 \times 10^{-5}$; $p53^{25,26,53,54/+}$, ** $P = 4.77 \times 10^{-7}$. **c,** Whole-mount images of a $p53^{25,26/+}$ embryo (right) at E9.5, displaying developmental delay (top) and neural tube defects, including exencephaly and kinked spine (bottom), compared with a littermate control (left). Original magnification, $\times 50$. **d,** $p53^{25,26,53,54/-}$ mice displayed a shorter lifespan (median lifespan, 128 days; $n = 8$) than wild-type mice (median lifespan, 774 days) and a similar lifespan to $p53^{+/-}$ mice (median lifespan, 143 days), further indicating that the $p53^{25,26,53,54}$ allele itself behaves like a $p53$ -null allele; $P < 0.0001$ by Mantel-Cox statistical analysis comparing wild-type and $p53^{25,26,53,54/-}$ mice.

a

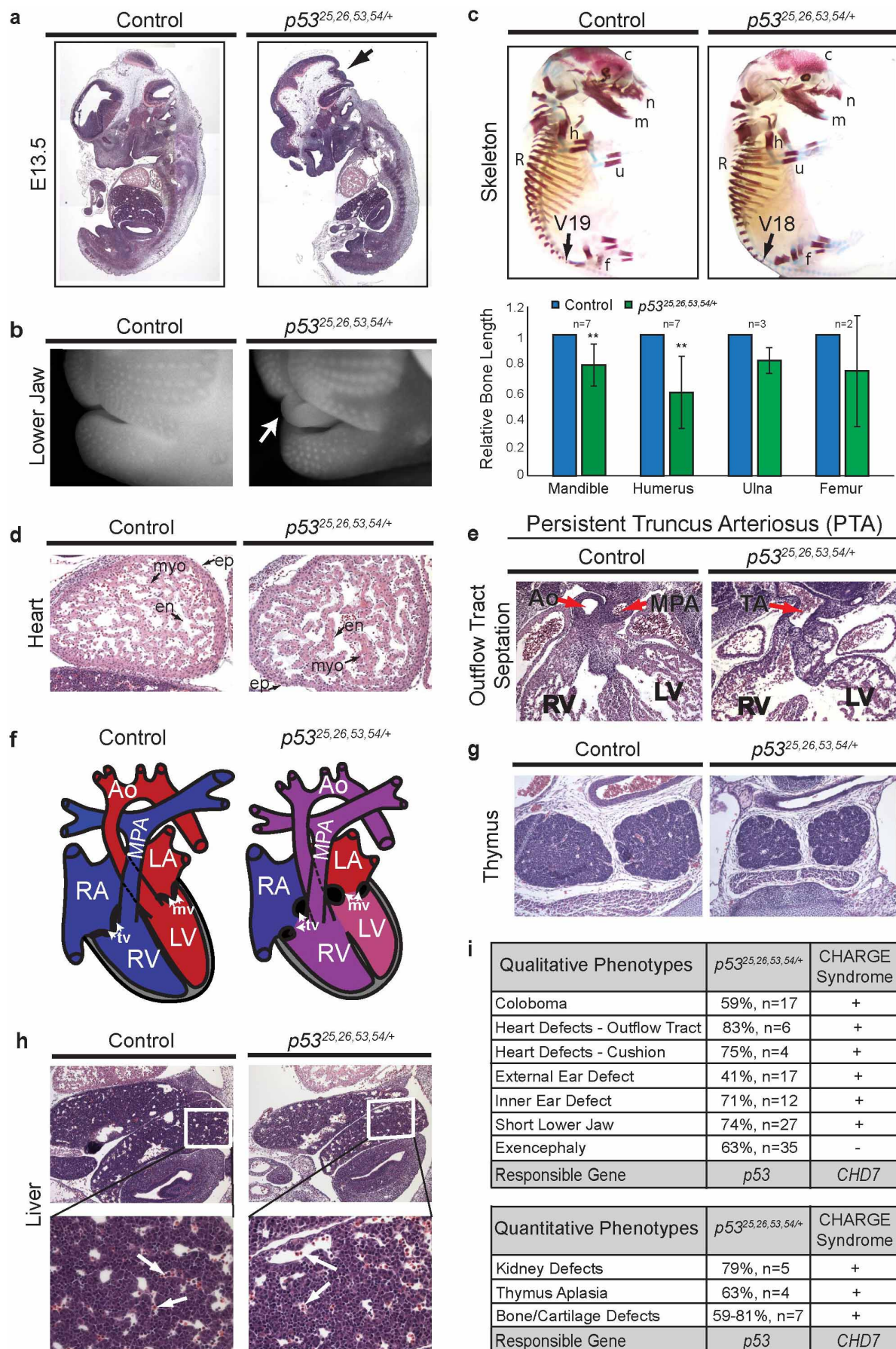
Figure	Genotype
Figure 1b	<i>p53</i> ^{+/+} ; <i>CMV-Cre</i>
Figure 1c	<i>p53</i> ^{+/+} ; <i>CMV-Cre</i>
Figure 1d	<i>p53</i> ^{+/+} ; <i>CMV-Cre</i>
Figure 1e	<i>p53</i> ^{+/+}
Figure 1f	<i>p53</i> ^{+/-} ; <i>CMV-Cre</i>
Figure 1g	<i>p53</i> ^{+/+} ; <i>CMV-Cre</i>
Figure 2a	<i>p53</i> ^{LSL-25,26,53,54/+}
Figure 2b	<i>p53</i> ^{+/+}
Figure 2c	<i>p53</i> ^{+/+}
Figure 2e	<i>p53</i> ^{+/+} ; <i>CMV-Cre</i>
Figure 2f	<i>p53</i> ^{+/+}
Figure 4f	<i>Chd7</i> ^{+/+} <i>p53</i> ^{+/-}
Extended Data Figure 2c	<i>p53</i> ^{+/+}
Extended Data Figure 4a	<i>p53</i> ^{+/+} ; <i>CMV-Cre</i>
Extended Data Figure 4b	<i>p53</i> ^{+/+} ; <i>CMV-Cre</i>
Extended Data Figure 4c	<i>p53</i> ^{+/+} ; <i>CMV-Cre</i>
Extended Data Figure 4d	<i>p53</i> ^{+/+} ; <i>CMV-Cre</i>
Extended Data Figure 4e	<i>p53</i> ^{+/+}
Extended Data Figure 4g	<i>p53</i> ^{+/+} ; <i>CMV-Cre</i>
Extended Data Figure 4h	<i>p53</i> ^{+/-} ; <i>CMV-Cre</i>
Extended Data Figure 5a	<i>p53</i> ^{+/-}
Extended Data Figure 5b	<i>p53</i> ^{+/-}
Extended Data Figure 5c	<i>p53</i> ^{+/-}
Extended Data Figure 5d	<i>p53</i> ^{+/-}
Extended Data Figure 5e	<i>p53</i> ^{+/+}
Extended Data Figure 5f	<i>p53</i> ^{+/-}
Extended Data Figure 5g	<i>p53</i> ^{+/+}
Extended Data Figure 5h	<i>p53</i> ^{+/-}
Extended Data Figure 6a	<i>p53</i> ^{+/+}
Extended Data Figure 6b	<i>p53</i> ^{LSL-25,26,53,54/+}
Extended Data Figure 6c	<i>p53</i> ^{+/+} ; <i>CMV-Cre</i>
Extended Data Figure 6d	<i>p53</i> ^{LSL-25,26,53,54/+}
Extended Data Figure 6e	<i>p53</i> ^{+/+} ; <i>CMV-Cre</i>
Extended Data Figure 6f	<i>p53</i> ^{LSL-25,26,53,54/+}

b

Phenotype	Male	Female
Coloboma	11	7
Short Jaw	10	10
Outer Ear	3	5
Exencephaly	10	6

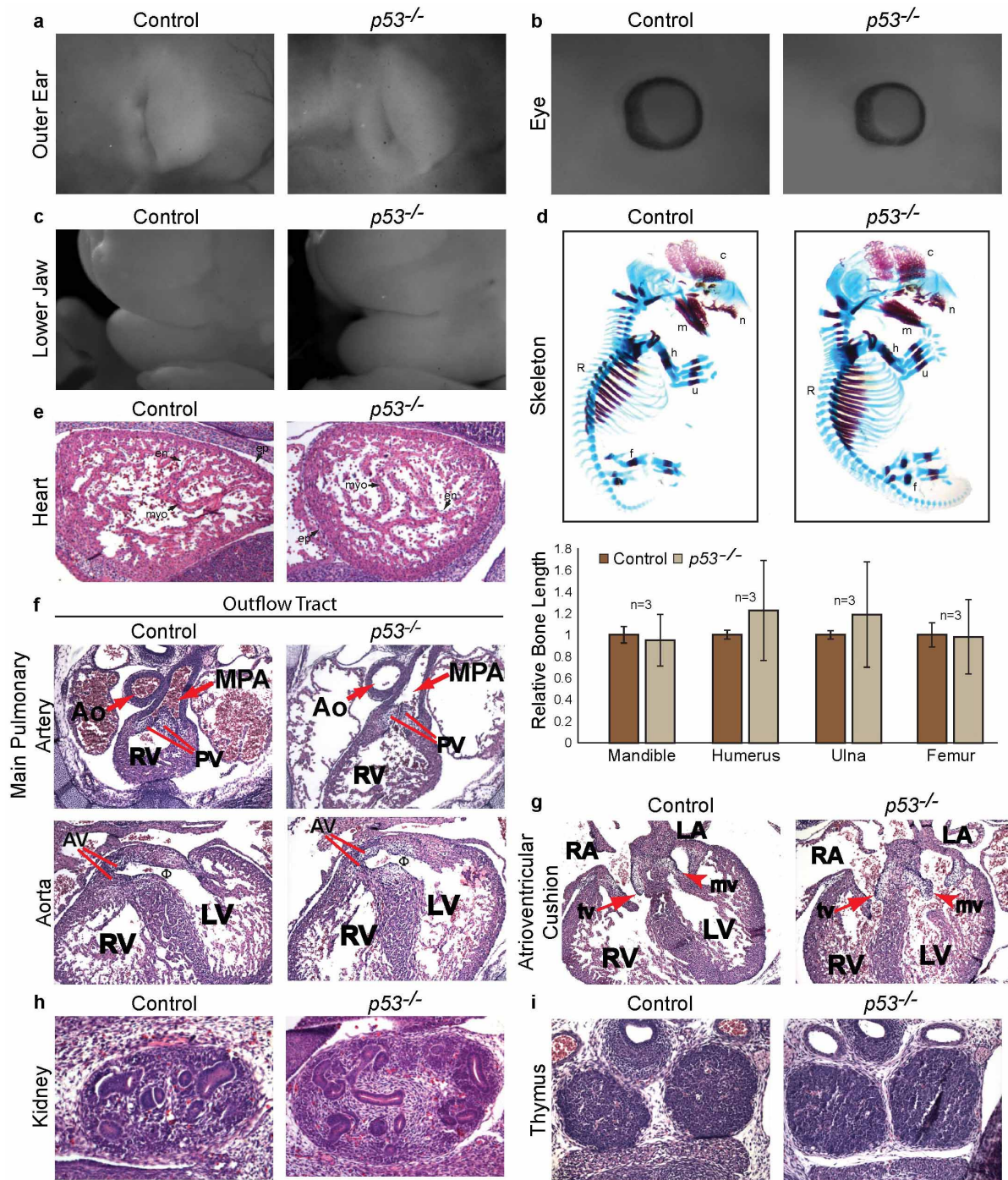
Extended Data Figure 3 | Genotypes of the control embryos in the figures and the genders associated with phenotypes. **a**, Table identifying the genotypes of the control embryos shown for each analysis. **b**, Table showing the

number of male and female *p53*^{25,26,53,54/+} embryos observed for the indicated phenotypes, as assessed by *Zfy* PCR. The phenotypes are well represented in both sexes.



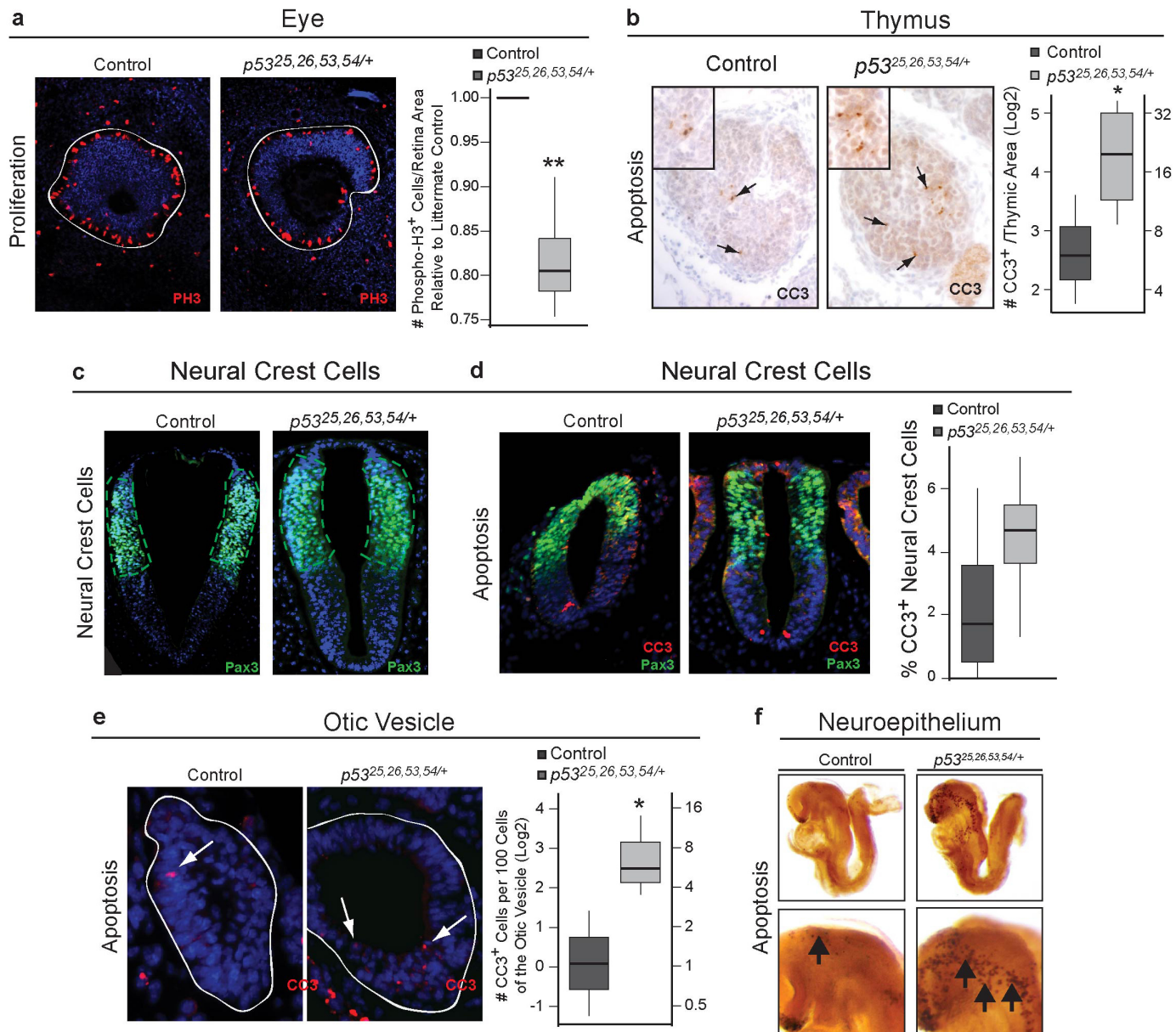
Extended Data Figure 4 | $p53^{25,26,53,54/+}$ embryos exhibit additional features of CHARGE syndrome. **a**, Haematoxylin and eosin stained sections of E12.5 control (left) and $p53^{25,26,53,54/+}$ embryos (right). Examination confirmed neural tube closure defects (arrow). Original magnification, $\times 10$. **b**, Close-up image of an ultraviolet-radiation-illuminated, ethidium bromide stained E15.5 $p53^{25,26,53,54/+}$ embryo (right), highlighting the short lower jaw phenotype with protruding tongue (arrow) compared with control littermates (left). Seventy-four per cent of $p53^{25,26,53,54/+}$ embryos ($n = 27$) exhibited a short lower jaw. Cleft lip is not shown. Original magnification, $\times 40$. **c**, Top, Alizarin red (bone) and Alcian blue (cartilage) stained whole mount of an E15.0 $p53^{25,26,53,54/+}$ embryo (right), showing reduced bone density in the cranium (c), nasal cavity (n), shorter ulna (u), humerus (h), mandible (m) and femur (f), as well as reduced bone formation in the ribs (R), where fewer vertebrae are undergoing ossification than in littermate controls (left). The number of vertebrae with bone formation was 19 in the controls (arrow; V19) and 18 in $p53^{25,26,53,54/+}$ embryos (arrow; V18). The severity of the bone and cartilage defects was variable, with the most severe defects evident in embryos with exencephaly and severe craniofacial defects ($n = 7$). Bottom, Quantification of bone lengths shown as percentage of E14.5–15.0 littermate controls. The bone lengths of the mandible, humerus, ulna and femur were measured using the ruler function in Adobe Photoshop on images acquired at $6.3\times$ magnification. Only litters with detectable bone formation in $p53^{25,26,53,54/+}$ embryos were included in bone length analyses: Student's *t*-test; **, $P = 0.008$ (mandible); **, $P = 0.005$ (humerus). **d**, Representative images of haematoxylin and eosin stained sagittal sections of E12.5 control hearts (left) and $p53^{25,26,53,54/+}$ hearts (right), showing all three cardiac cell types in both genotypes. en, endocardium; ep, epicardium; myo, myocardium (arrows). Original magnification, $\times 200$. **e**, A haematoxylin and eosin stained E12.5 $p53^{25,26,53,54/+}$ heart exhibiting persistent truncus arteriosus (PTA) (33%, $n = 6$). The cardiac outflow tract in the control embryo (left) is septated into the aorta (Ao) and main pulmonary

artery (MPA), whereas the cardiac outflow tract (truncus arteriosus (TA)) in the $p53^{25,26,53,54/+}$ embryo (right) remains unseptated, resulting in PTA. Original magnification, $\times 100$. **f**, Illustration of a control heart (left) and a $p53^{25,26,53,54/+}$ embryo heart (right), highlighting double outlet right ventricle (DORV) and atrioventricular cushion defects. Both the aorta (Ao) and the MPA flow out of the right ventricle (RV), resulting in mixed oxygenated and deoxygenated blood in the systemic circulation when combined with concurrent ventricular septal defects (VSDs). The atrioventricular cushions remain bulbous and fail to elongate into mature valve leaflets. Red denotes oxygenated blood; blue denotes deoxygenated blood; and purple (pink) denotes mixed oxygenated and deoxygenated blood. mv, mitral valve; tv, tricuspid valve. Original magnification, $\times 100$. **g**, Representative haematoxylin and eosin stained transverse section of thymus from a $p53^{25,26,53,54/+}$ E15.5 embryo (right) reveals a smaller thymus than in littermate controls (left) (63% of controls; $n = 4$). Original magnification, $\times 200$. **h**, Representative haematoxylin and eosin analysis of liver sections from E12.5 controls (left) and $p53^{25,26,53,54/+}$ embryos (right), showing normal liver architecture in both genotypes (top). High magnification image (bottom) of the region of the liver that is outlined by the white box in the top panel shows the presence of nucleated erythrocytes (arrows), indicating proper haematopoiesis. Original magnification top, $\times 100$; bottom, $\times 400$. **i**, Top, Table summarizing the incidence (%) and sample size (n) of phenotypes assessed qualitatively in $p53^{25,26,53,54/+}$ embryos. The occurrence of these phenotypes in CHARGE syndrome is also indicated (+, present; –, absent). Bottom, Table summarizing the phenotypes assessed quantitatively in $p53^{25,26,53,54/+}$ embryos relative to controls, shown as the percentage average size of the control (%), with sample size (n) also indicated. The occurrence of these phenotypes in CHARGE syndrome is also shown (+, present). A detailed description of the bone and cartilage defects is provided in c.



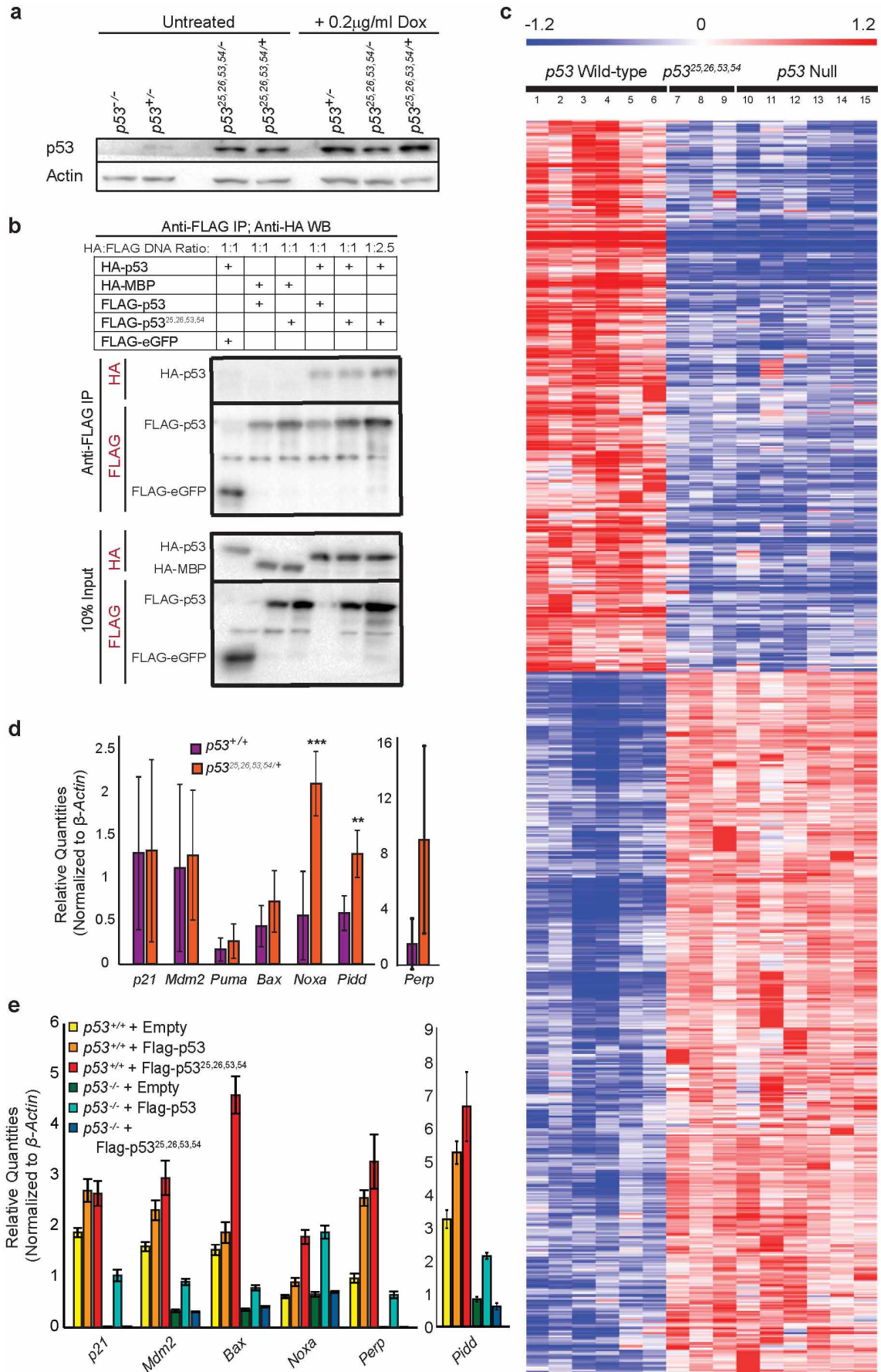
Extended Data Figure 5 | $p53^{-/-}$ embryos do not exhibit characteristics of CHARGE syndrome. **a**, Whole-mount image of the external ear of an E15.5 $p53^{-/-}$ embryo (right) and a control embryo (left), showing normal ear pinna development. Original magnification, $\times 100$. **b**, Whole-mount image of an E13.5 $p53^{-/-}$ embryo (right) and a control embryo (left), showing normal retinal development and no evidence of coloboma. Original magnification, $\times 100$. **c**, Whole-mount image of an E15.5 $p53^{-/-}$ embryo (right) and a control embryo (left), showing normal lower jaw development. Original magnification, $\times 40$. **d**, Top, Alizarin red (bone) and Alcian blue (cartilage) stained whole-mount E14.5 $p53^{-/-}$ embryo (right), showing normal long bone formation of the ulna (u), humerus (h), mandible (m), and femur (f) relative to littermate controls (left). Bottom, Quantification of the bone lengths shown as a percentage of E14.5 littermate controls ($n = 3$). Original magnification, $\times 6.3$. **e**, Representative images of haematoxylin and eosin stained sagittal sections of E13.5 control hearts (left) and $p53^{-/-}$ hearts (right), showing all three cardiac cell types in both genotypes. en, endocardium; ep, epicardium; myo, myocardium (arrows). Original magnification, $\times 200$. **f**, Analysis of

haematoxylin and eosin stained transverse sections of E13.5 $p53^{-/-}$ and control hearts revealing normal outflow tract development. Top, The MPA and Ao are fully septated, and the MPA connects to the right ventricle (RV) in $p53^{-/-}$ hearts. Bottom, The Ao connects to the left ventricle (LV). The symbol Φ denotes that the ventricular outflow tract connects the LV and Ao. AV, aortic valve; PV, pulmonary valve. Original magnification, $\times 100$. **g**, Analysis of transverse sections of haematoxylin and eosin stained E13.5 $p53^{-/-}$ hearts (right) reveals normal atrioventricular cushions, which have undergone remodelling to form mature, elongated mitral valves (mv; arrowhead) and tricuspid valves (tv; arrow) similar to in control hearts (left). LA, left atrium; RA, right atrium. Original magnification, $\times 100$. **h**, Haematoxylin and eosin stained sagittal section of kidney from $p53^{-/-}$ (right) and control (left) embryos, showing normal renal size and development. Original magnification, $\times 200$. **i**, Haematoxylin and eosin stained transverse section of thymus in a $p53^{-/-}$ E13.5 embryo (right) reveals a similar thymus size to that in a littermate control (left). Original magnification, $\times 200$.



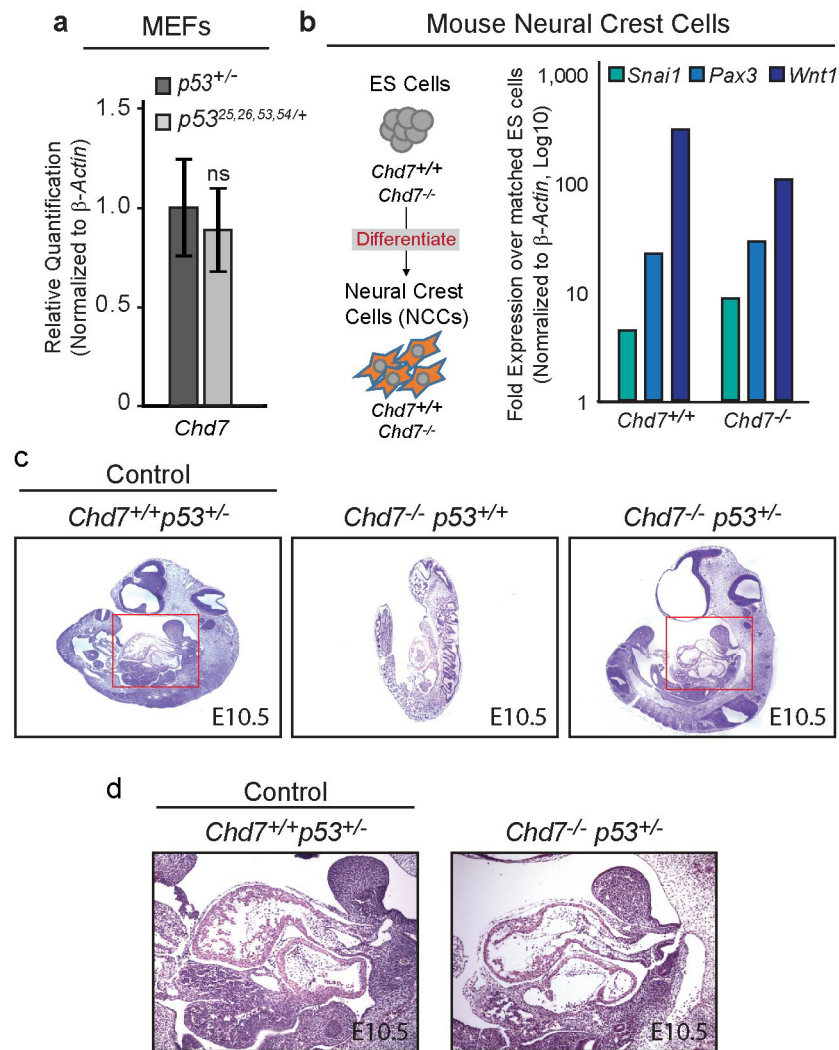
Extended Data Figure 6 | $p53^{25,26,53,54/+}$ embryo tissues display increased apoptosis and decreased proliferation. **a**, Left, Immunofluorescence for phospho-histone H3 (red) in the retina of E13.5 control and $p53^{25,26,53,54/+}$ embryos. Right, Quantification of phospho-H3-positive cells per retina area relative to littermate controls. **, $P = 0.006$ by one-tailed Welch's t -test ($n = 4$). Original magnification, $\times 200$. **b**, Left, Immunohistochemistry for cleaved caspase 3 (CC3) in thymi of control (left) and $p53^{25,26,53,54/+}$ (right) embryos. Inset, Magnified image of CC3-positive region. Right, Quantification of CC3-positive cells per thymic area. *, $P = 0.02$ by one-tailed Student's t -test ($n = 4$). Original magnification, $\times 400$. **c**, Immunofluorescence for PAX3 (green) in NCCs of E9.5 control and $p53^{25,26,53,54/+}$ embryos was used to identify NCCs in Fig. 2f. Original magnification, $\times 200$. **d**, Left, Immunofluorescence for CC3 (red) and PAX3 (green) in NCCs of E9.5 control

and $p53^{25,26,53,54/+}$ embryos. $p53^{25,26,53,54/+}$ embryos have more apoptotic (red) NCCs, as determined by PAX3-positive staining (green), than littermate controls. Right, Quantification of CC3-positive cells per total NCC number. $P = 0.14$ by one-tailed Student's t -test ($n = 4$). Original magnification, $\times 200$. **e**, Left, Immunofluorescence for CC3 (red) in the otic vesicles of E9.5 control and $p53^{25,26,53,54/+}$ embryos. Right, Quantification of CC3-positive cells per total cell number. *, $P = 0.03$ by one-tailed Student's t -test ($n = 3$). Original magnification, $\times 200$. **f**, CC3 staining in whole-mount E8.5 control and $p53^{25,26,53,54/+}$ embryos reveals enhanced apoptosis in the neuroepithelium of $p53^{25,26,53,54/+}$ embryos (right) but not in controls (left). The close-up shows a magnification of the caudal neuroepithelium (bottom). Arrows indicate CC3-positive regions. Original magnification top, $\times 50$.



Extended Data Figure 7 | $p53^{25,26,53,54}$ is transactivation dead but augments wild-type p53 activity. **a**, Western blot analysis of p53 protein levels in untreated or doxorubicin-treated ($0.2 \mu\text{g ml}^{-1}$ Dox) $p53^{-/-}$, $p53^{+/-}$, $p53^{25,26,53,54/-}$ and $p53^{25,26,53,54/+}$ MEFs. β -Actin served as a loading control. **b**, Western blot analysis of anti-Flag immunoprecipitation from $p53^{-/-}$ MEFs transiently overexpressing HA-p53 and Flag-p53 or Flag- $p53^{25,26,53,54}$. HA-MBP and Flag-eGFP were used as negative controls. Immunoprecipitated protein and 10% input were probed with either anti-HA or anti-Flag antibodies. (The microgram ratio of HA-p53 to Flag-p53 or Flag- $p53^{25,26,53,54}$ plasmid DNA was 1:1 or 1:2.5, as indicated above the blot (see Fig. 3b). **c**, Heat map examining the transactivation capacity of $p53^{25,26,53,54}$ on p53-dependent genes identified by microarray analysis through comparison of six *HrasV12*; WT *p53* MEF lines with six *HrasV12*; *p53*-null MEF lines, as previously

described³. Three independent *HrasV12*; $p53^{25,26,53,54/25,26,53,54}$ MEF lines were analysed, and the gene expression profiles were indistinguishable from those of *HrasV12*; *p53*-null cells. The numbered columns indicate independent MEF lines. Blue denotes repressed genes, and red denotes induced genes. **d**, qRT-PCR analysis of p53 target gene expression in untreated MEFs derived from $p53^{+/-}$ and $p53^{25,26,53,54/+}$ E13.5 embryos. Graphs indicate the mean \pm s.d. from four independent MEF lines after normalization to β -actin gene expression. **, $P < 0.01$; ***, $P < 0.005$; Student's *t*-test. **e**, qRT-PCR analysis of p53 target gene expression in $p53^{+/-}$ and $p53^{-/-}$ MEFs stably transduced with empty vector, Flag-p53 or Flag- $p53^{25,26,53,54}$. The representative gene expression from one experiment is shown: mean \pm s.d. of technical triplicates after normalization to β -actin gene expression. The experiment was performed in duplicate.



Extended Data Figure 8 | p53 heterozygosity partially rescues *Chd7*-null embryos. **a**, qRT-PCR analysis of *Chd7* expression in untreated MEFs derived from E13.5 $p53^{+/-}$ and $p53^{25,26,53,54/+}$ embryos. The graphs indicate the mean \pm s.d. from four independent MEF lines after normalization to β -actin gene expression. ns, not significant. **b**, Left, Schematic of NCC differentiation. Right, Representative qRT-PCR analysis of NCC markers in NCC-like cells differentiated from $Chd7^{+/+}$ and $Chd7^{-/-}$ (*whi/whi*) mouse embryonic

stem cells normalized to β -actin gene expression and relative to matched embryonic stem cells. **c**, Haematoxylin and eosin stained E10.5 $Chd7^{+/+}p53^{+/-}$ (control), $Chd7^{-/-}p53^{+/+}$ and $Chd7^{-/-}p53^{+/-}$ embryos. The $Chd7^{-/-}p53^{+/+}$ embryo shown is necrotic, as evidenced by cellular autolysis. Original magnification, $\times 32$. **d**, Close-up image of heart region denoted by red box in panel **c**, in E10.5 $Chd7^{+/+}p53^{+/-}$ (control) and $Chd7^{-/-}p53^{+/-}$ embryos. Original magnification, $\times 100$.

Extended Data Table 1 | $p53^{25,26,53,54/+}$ mice do not survive to weaning

$p53^{+/+};CMV-Cre$ X $p53^{LSL-X/+}$			
Genotype:	$p53^{LSL-wt}$	$p53^{LSL-53,54}$	$p53^{LSL-25,26,53,54}$
# Observed Progeny (# Expected Progeny) at P21			
$p53^{+/+}$	18 (15)	10 (12)	19 (41.75)
$p53^{+/+} Cre^{+}$	21 (15)	14 (12)	68 (41.75)
$p53^{LSL-X/+}$	11 (15)	9 (12)	50 (41.75)
$p53^{X/+} Cre^{+}$	10 (15)	15 (12)	0* (41.75)
	n=60 p=0.086	n=48 p=0.876	n=167 *p=7.65E-18

Crosses of $p53^{LSL-wt/+}$, $p53^{LSL-53,54/+}$ or $p53^{LSL-25,26,53,54/+}$ with $p53^{+/+};CMV-Cre$ mice yield the expected numbers of $p53^{wt/+}$ and $p53^{53,54/+}$ pups but not $p53^{25,26,53,54/+}$ pups at postnatal day 21 (P21). X denotes wild-type or mutant allele. The observed numbers of live and dead pups compared with the expected numbers of live pups are indicated: [Observed (Expected)]. $p53^{53,54}$ mutations in the second TAD did not perturb the transcriptional activity of p53 at the genome-wide level and provide a control for the mutation of only these residues. The presence of viable $p53^{wt/+}$ and $p53^{53,54/+}$ pups suggests that re-expression of p53 in early development using this system does not cause developmental defects. The binomial distribution statistical test shows that the absence of $p53^{25,26,53,54/+}$ pups at P21 is significant. *, $P = 7.65 \times 10^{-18}$.

Extended Data Table 2 | $p53^{25,26,53,54/+}$ embryo survival and incidence of exencephaly in live embryos

p53 ^{+/+} ; CMV-Cre X p53 ^{LSL-25,26,53,54/+}						
Genotype:		Embryonic Age:				
		E12.5	E13.5	E14.5	E15.5	E18.5
# Observed Progeny (# Expected Progeny)						
p53 ^{+/+}	Live	32 (28)	37 (45)	21 (20)	15 (19)	12 (14)
	Dead	1	3	2	0	0
	Exencephaly	0	0	0	0	0
p53 ^{+/+} Cre ⁺	Live	27 (28)	59 (45)	20 (20)	17 (19)	20 (14)
	Dead	4	12	3	1	5
	Exencephaly	0	0	0	1	0
p53 ^{LSL-25,26,53,54/+}	Live	26 (28)	38 (45)	18 (20)	24 (19)	9 (14)
	Dead	3	3	1	0	0
	Exencephaly	0	0	0	0	0
p53 ^{25,26,53,54/+} Cre ⁺	Live	16 (19)	23* (37)	7 (14)	1** (16)	2* (7)
	Dead	3	8	11	6	9
	Exencephaly	14	8	2	1	1
p53 ^{LSL-25,26,53,54/+} Cre ⁺ No or incomplete recombination	Live	9	8	6	3	7
	Dead	1	1	1	0	1
	Exencephaly	0	2	0	0	0
		n=122 p=0.290	n=192 *p=0.019	n=90 p=0.053	n=67 **p=1.24E-05	n=65 *p=0.047

Survival of embryos at E12.5, E13.5, E14.5, E15.5 and E18.5 from timed matings between $p53^{LSL-25,26,53,54/+}$ and $p53^{+/+}; CMV-Cre$ mice. The numbers of live mice of each genotype were identified by the presence of a heartbeat. The observed numbers of live and dead pups compared with the expected numbers of live pups are indicated: [Observed (Expected)]. For the $p53^{25,26,53,54/+}; Cre$ genotype, the number of expected live pups was fewer than for the other genotypes owing to some incidence of incomplete recombination of the LSL element, as detailed below. The number of live embryos exhibiting exencephaly observed at the time of dissection is also indicated (63% over all time points). Females and males, as determined by Zfy PCR, exhibited exencephaly with the same frequency. The genotypes of $p53^{25,26,53,54/+}$ embryos carrying a $CMV-Cre$ transgene lack the LSL designation because the $lox-Stop-lox$ element has been deleted from the genome. As reported³¹, we observed mosaic Cre activity in some embryos, reflected by incomplete or no $lox-Stop-lox$ deletion in PCR analysis of yolk sac DNA. Embryos that showed little to no recombination of the $lox-Stop-lox$ allele were excluded from analyses. Statistical significance for the viability of embryos was assessed using binomial distribution analysis to account for embryos with little or no recombination of the $lox-Stop-lox$ allele. The binomial distribution statistical test showed that the absence of live $p53^{25,26,53,54/+}$ embryos at late-gestational time points is significant. E12.5, $P = 0.290$; E13.5, *, $P = 0.019$; E14.5, $P = 0.053$; E15.5, **, $P = 1.24 \times 10^{-5}$; E18.5, *, $P = 0.047$. Logistic regression to compare the viability of $CMV-Cre$; $p53^{25,26,53,54/+}$ embryos with and without complete recombination showed that the increased death of recombined $CMV-Cre$; $p53^{25,26,53,54/+}$ embryos is significant: $P < 0.0001$.

Protein competition switches the function of COP9 from self-renewal to differentiation

Lei Pan^{1,2*}, Su Wang^{1,3*}, Tinglin Lu^{4*}, Changjiang Weng¹, Xiaoqing Song¹, Joseph K. Park^{5,6}, Jin Sun⁴, Zhi-Hao Yang⁴, Junjing Yu^{1,2}, Hong Tang², Dennis M. McKearin^{5,6}, Daniel A. Chamovitz⁷, Jianquan Ni⁴ & Ting Xie^{1,3}

The balance between stem cell self-renewal and differentiation is controlled by intrinsic factors and niche signals^{1,2}. In the *Drosophila melanogaster* ovary, some intrinsic factors promote germline stem cell (GSC) self-renewal, whereas others stimulate differentiation³. However, it remains poorly understood how the balance between self-renewal and differentiation is controlled. Here we use *D. melanogaster* ovarian GSCs to demonstrate that the differentiation factor Bam controls the functional switch of the COP9 complex from self-renewal to differentiation via protein competition. The COP9 complex is composed of eight Csn subunits, Csn1–8, and removes Nedd8 modifications from target proteins^{4,5}. Genetic results indicated that the COP9 complex is required intrinsically for GSC self-renewal, whereas other Csn proteins, with the exception of Csn4, were also required for GSC progeny differentiation. Bam-mediated Csn4 sequestration from the COP9 complex via protein competition inactivated the self-renewing function of COP9 and allowed other Csn proteins to promote GSC differentiation. Therefore, this study reveals a protein-competition-based mechanism for controlling the balance between stem cell self-renewal and differentiation. Because numerous self-renewal factors are ubiquitously expressed throughout the stem cell lineage in various systems, protein competition may function as an important mechanism for controlling the self-renewal-to-differentiation switch.

D. melanogaster ovarian GSCs are an attractive model for investigating stem-cell–niche interactions at the molecular and cellular level³. In the germarium of the *D. melanogaster* ovary, two or three GSCs can be easily identified by their direct contact with cap cells and a spherical fusome (also known as a spectrosome), which is apically anchored^{6–8}. Bam controls GSC daughter differentiation by forming a protein complex with Bgcn^{9–12}. Niche-activated Bmp signalling is necessary and sufficient for repressing *bam* gene transcription in GSCs, thereby maintaining GSC self-renewal^{7,13,14}. Immediate differentiating GSC daughters (also known as cystoblasts) also contain a spherical spectrosome and begin *bam* transcription^{6,15}. These cells can further divide without cytokinesis to form 2-cell, 4-cell, 8-cell and 16-cell cysts, which harbour a branched fusome¹⁶. Although Bam is a master regulator of GSC differentiation¹⁰, it remains largely unclear whether and how Bam inactivates self-renewal factors in differentiating GSC progeny. In this study, we show that Bam converts the function of the COP9 complex from self-renewal to differentiation by sequestering Csn4.

In the yeast two-hybrid screen described in our previous study¹¹, the carboxy-terminal 121 amino acid region of Csn4 was identified to interact with Bam (Extended Data Fig. 1a). In yeast (*Saccharomyces cerevisiae*) and *D. melanogaster* S2 cells, Bam also interacts with full-length Csn4 (Extended Data Fig. 1b–d). Using yeast two-hybrid interaction experiments, we found that the 151–350 amino acid central domain of Bam interacts with Csn4 (Extended Data Fig. 1b, c). Additionally, *nos*-GAL4-driven germline-specific expression of haemagglutinin (HA)-tagged

Csn4 pulled down endogenous Bam *in vivo*, but germline-expressed Flag-tagged Csn5 failed to pull down Bam (Fig. 1a). Finally, we generated P[acman]-based bacterial artificial chromosome (BAC) transgenes¹⁷, in which a Flag tag was added to the coding region of *Csn4* and *Csn5* (Extended Data Fig. 2), and we showed that the Csn4 and Csn5 proteins exhibit similar ubiquitous expression patterns in germ cells and are co-expressed with Bam in mitotic cysts (Fig. 1b–d). These results demonstrate that Bam interacts with Csn4 to form a protein complex *in vivo*.

In our genetic screen, *Csn4* was identified as a dominant suppressor of the germ cell differentiation defect in the hypomorphic *bam^Z/bam^{Δ86}* mutant. Immunostaining for Hts labels spectrosomes and fusomes, allowing the identification of GSCs, cystoblasts and differentiated cysts¹⁶. In the *bam^Z/bam^{Δ86}* mutant ovaries, the germaria and most of the egg chambers were filled with undifferentiated cystoblasts (Fig. 1e). By contrast, inactivation of one copy of the *Csn4* gene using a deletion allele (*Csn4ⁿ*) or a P element insertion allele (*Csn4^k*)¹⁸ dramatically reduced the number of undifferentiated cystoblasts and dramatically increased the number of differentiated cysts containing a branched fusome in the *bam^Z/bam^{Δ86}* germaria and the number of normal egg chambers per ovariole (Fig. 1f–h and Extended Data Fig. 3a, b, b'). These results suggest that Csn4 and Bam have opposite functions in regulating germ cell differentiation.

To determine whether *Csn4*, *Csn5* and *Nedd8* are intrinsically required to maintain GSCs, we used FLP-mediated FRT recombination to inactivate the functions of Csn4, Csn5 and Nedd8 in GSCs by constructing control and mutant *Csn4*, *Csn5* and *Nedd8* GSC clones marked by the absence of *Ubi-GFP* (green fluorescent protein (GFP) controlled by the *Ubiquitin* promoter) or *armadillo-lacZ* as in our previous studies^{7,19}. *Csn5ⁿ* and *Nedd8^{ΔN015}* represent a deletion allele and a hypomorphic allele, respectively^{18,20,21}. In contrast to the marked control GSCs, which exhibited a slow natural turnover during the first 3-week period after clone induction (ACI) (Fig. 2a, b), the marked mutant *Csn4^k*, *Csn4ⁿ*, *Csn5ⁿ* and *Nedd8^{ΔN015}* GSCs were lost much faster, and most of them were therefore lost from the niche at 3 weeks ACI (Fig. 2b, c). This finding is consistent with the previous finding that *Nedd8* and *Csn* mutants exhibit similar phenotypes in *D. melanogaster* and plants^{22,23}. Then, we used the combined Gal4–UAS and FLP-Out system to knock down the functions of the other *Csn* genes in adult GSCs by RNA interference and then examined the GSC numbers in germline-specific knock-down (GSKD) germaria 3 days (3 d), 7 d and 14 d after heat shock (AHS)²⁴ (Extended Data Fig. 4a). In contrast to control *Csn1a^{GSKD}* (a pseudo gene) and *GFP^{GSKD}* germaria, which maintained two or three GSCs 3 d, 7 d and 14 d AHS, *Csn1b^{GSKD}*, *Csn2^{GSKD}*, *Csn3^{GSKD}*, *Csn6^{GSKD}* and *Csn7^{GSKD}* germaria showed a significant GSK loss, and most of them contained only one GSC on average at 7 d and 14 d AHS (Fig. 2d, e and Extended Data Fig. 4b–g). Noticeably, *Csn8* knock down yielded no GSK loss

¹Stowers Institute for Medical Research, 1000 East 50th Street, Kansas City, Missouri 64110, USA. ²Chinese Academy of Sciences Key Laboratory of Infection and Immunity, Institute of Biophysics, 15 Da Tun Road, Beijing 100101, China. ³Department of Cell Biology and Anatomy, University of Kansas School of Medicine, 3901 Rainbow Boulevard, Kansas City, Kansas 66160, USA. ⁴Center for Life Sciences, School of Medicine, Tsinghua University, Beijing 100084, China. ⁵Department of Molecular Biology and Graduate School of Biomedical Sciences, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9148, USA. ⁶Howard Hughes Medical Institute, Chevy Chase, Maryland 20815-6789, USA. ⁷Department of Plant Sciences, Tel Aviv University, Tel Aviv 69978, Israel.

*These authors contributed equally to this work.

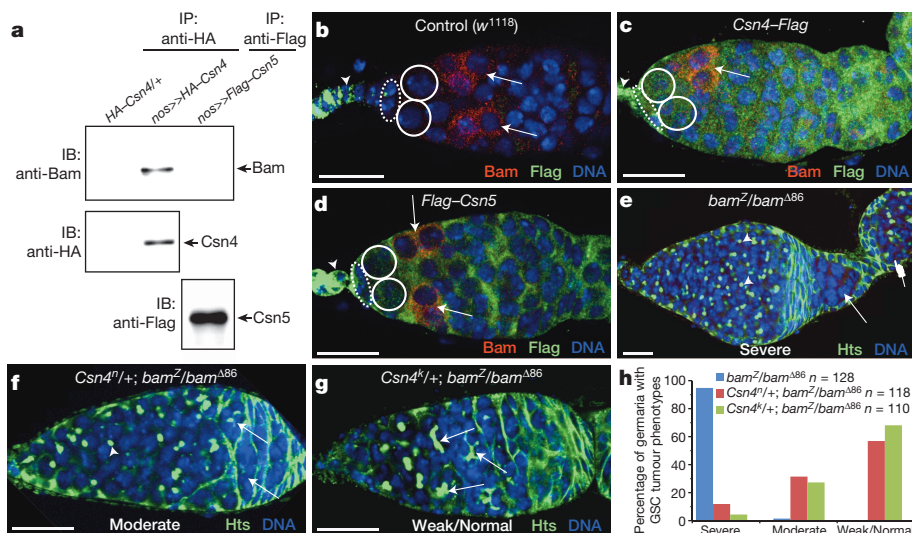


Figure 1 | Csn4 physically interacts with Bam and antagonizes its function during GSC development. **a**, HA-Csn4, but not Flag-Csn5, pulled down endogenous Bam *in vivo*, as determined by co-immunoprecipitation. IB, immunoblot; IP, immunoprecipitation. **b–d**, Csn4-Flag (**c**) and Flag-Csn5 (**d**) transgenes showed ubiquitous expression in GSCs (solid circles) and in their progeny and cap cells (dashed ovals) compared with the control (**b**). Csn4 and Csn5 were co-expressed with Bam in mitotic cysts (arrows). Arrowheads indicate non-specific staining. **e–h**, Heterozygous Csn4^f (**f**) and Csn4^k (**g**) mutations suppressed the differentiation defect of the hypomorphic *bam* mutant (**e**) based on spectrosomes (arrowheads) and branched fusomes (arrows). **h**, Quantification of germaria with GSC tumour phenotypes in severe (**e**), moderate (**f**) and weak (**g**) phenotypes. Scale bars, 10 μ m.

phenotype, which might have been due to a low knock-down efficiency (Fig. 2e and Extended Data Fig. 4g). Taken together, these results indicate that the whole COP9 complex is likely to be intrinsically required for GSC maintenance.

TUNEL (terminal deoxynucleotidyl transferase dUTP nick end labelling) results showed that marked mutant Csn4^k, Csn4ⁿ and Csn5ⁿ GSCs and cysts exhibited no dramatic increase in apoptosis in comparison with marked control GSCs and cysts (Extended Data Fig. 5a–d). Interestingly, marked mutant Csn4ⁿ or Csn5ⁿ GSCs also produced fewer differentiated cysts than did marked control GSCs (Fig. 2a, c). Additionally, fewer Csn4 and Csn5 mutant GSCs than control GSCs were positive for 5-bromodeoxyuridine (BrdU) labelling and phosphorylated histone 3 (p-H3, a mitotic marker) (Extended Data Fig. 5e–h). Although COP9 regulates different signalling pathways in organisms ranging from plants to animals^{4,5}, Csn4 and Csn5 are dispensable for Bmp signalling and E-cadherin-mediated cell adhesion in GSCs (Extended Data Fig. 6). These results suggest that Csn4 and Csn5 are intrinsically required to control GSC self-renewal and proliferation independently of Bmp signalling and E-cadherin.

To understand how Csn4 mutations suppress the *bam* differentiation defect, we examined the protein interactions among Bam, Bgcn and Csn4. In yeast cells, Bgcn binds to the same central Bam domain (amino acids 151–350) as Csn4 does but has a weaker interaction with Bam (Extended Data Fig. 7a–d, d'). The presence of increasing concentrations of Csn4 gradually weakened the Bam–Bgcn interaction, whereas a truncated Csn4 protein lacking the Bam-interacting domain was unable to compete with Bgcn for Bam binding (Fig. 3a and Extended Data Fig. 7e). These results support the Csn4–Bam–Bgcn competition model, which predicts that Csn4 and Bam antagonize each other's function in the regulation of germ cell differentiation.

The numbers of cystoblasts and two-cell pairs (dividing cystoblasts and two-cell cysts) were then used to quantify germ cell differentiation defects because it is difficult to distinguish two-cell pairs that are dividing symmetrically to generate cystoblasts from normal two-cell cysts (Extended Data Fig. 7f). Interestingly, germ-cell-specific Csn4 overexpression in the *bam*^{Δ86} heterozygous germaria but not in wild-type germaria significantly increased the number of cystoblasts and two-cell pairs compared with that in *bam*^{Δ86} heterozygous germaria, but the

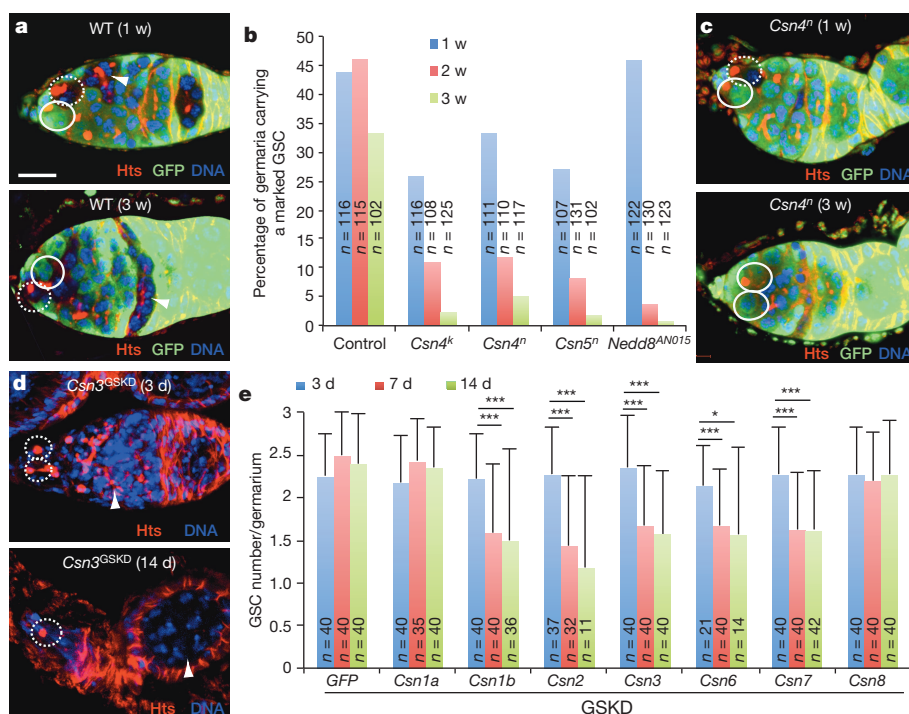


Figure 2 | COP9 is an intrinsic controller of GSC self-renewal and proliferation. **a–c**, In contrast to the marked control GSCs that are maintained at 3 weeks (3 w) ACI (**a**), the marked Csn4ⁿ mutant GSC (**c**) detected 1 w ACI (top) has been lost by 3 w ACI (bottom). The solid circles highlight unmarked GSCs (GFP⁺), whereas the dashed circles indicate marked GSCs (GFP[−]) or mutant GSCs. The arrowheads indicate marked cysts (**a**). The changes in the percentage of germaria carrying a marked control, Csn4, Csn5 or Nedd8 mutant GSC at 1, 2 or 3 w ACI are shown (**b**). **d**, Csn3^{GSKD} germaria contained two GSCs and one GSC at 3 days (3 d) and 14 d AHS, respectively. The arrowheads indicate differentiated germ cells. **e**, Csn1b, Csn2, Csn3, Csn6 and Csn7 (but not Csn8, GFP or Csn1a) knock-down germaria contained significantly fewer GSCs at 7 d and 14 d AHS than at 3 d. *, *P* < 0.05; ***, *P* < 0.001. Scale bars, 10 μ m.

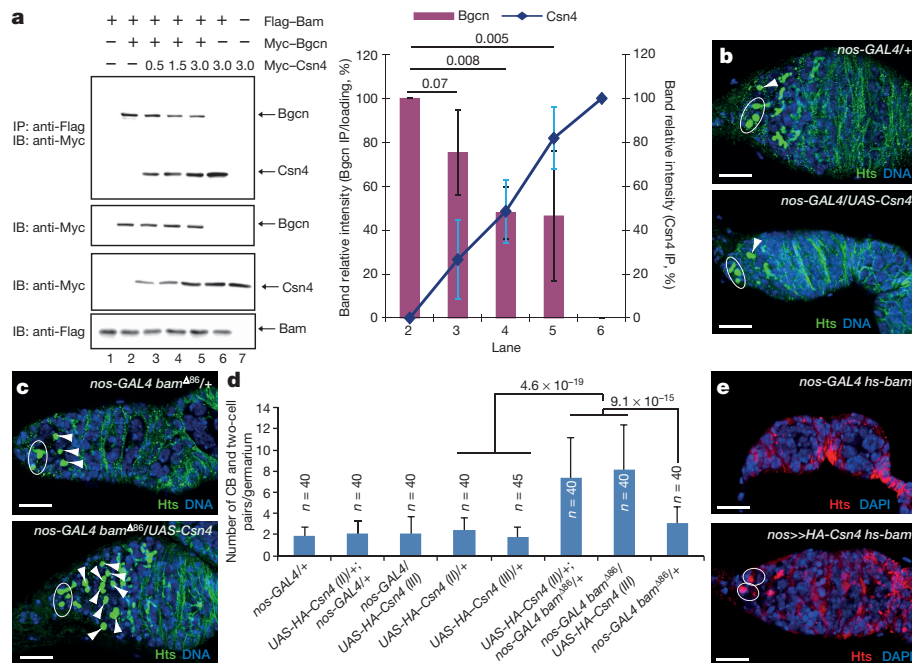


Figure 3 | Csn4 inhibits cystoblast differentiation by competing with Bgcn for binding to Bam. **a**, Csn4 outcompetes Bgcn for binding to Bam in a concentration-dependent manner in S2 cells, as determined by co-immunoprecipitation. Three independent experiments are represented in the histogram on the right, and *P* values are indicated. The numbers above the blot on the left indicate the amounts of DNA constructs in micrograms. **b–d**, The *nos-GAL4 bam^{Δ86}/UAS-Csn4* germarium (c, bottom) had more cystoblasts (CBs) and two-cell pairs (arrowheads) than the *nos-GAL4/+* (b, top), *nos-GAL4/UAS-Csn4* (b, bottom) and *nos-GAL4 bam^{Δ86}/+* (c) germaria. GSCs are highlighted by ovals. Quantification and *P* values are shown in **d**. Roman numerals on the *x* axis indicate sample sizes. **e**, The *nos-GAL4 hs-bam/+* germarium had lost all GSCs by 1 week AHS (top), but the *nos>>HA-Csn4 hs-bam* (*UASp-HA-Csn4/+; nos-GAL4 hs-bam/+*) germarium retained two GSCs (circles) 1 week AHS (bottom). Scale bars, 10 μ m.

germaria of both genotypes showed no significant differences in GSC numbers, indicating that excess Bam in differentiating GSC progeny suppresses the antagonizing function of Csn4 and thereby ensures their differentiation (Fig. 3b–d and Extended Data Fig. 7g). Ectopic Bam expression in GSCs via *hs-bam* induced the rapid differentiation of GSCs and their departure from the niche¹⁰. Indeed, a 1-h Bam induction

induced 83–95% of the germaria to completely lose GSCs (Fig. 3e and Extended Data Fig. 7h). However, germline-specific *Csn4* overexpression allowed 65% of the germaria to retain GSCs, indicating that Csn4 can also antagonize the differentiation-promoting function of ectopic Bam in GSCs (Fig. 3e, bottom, and Extended Data Fig. 7h). Taken together, these results suggest that differentiating GSC progeny express

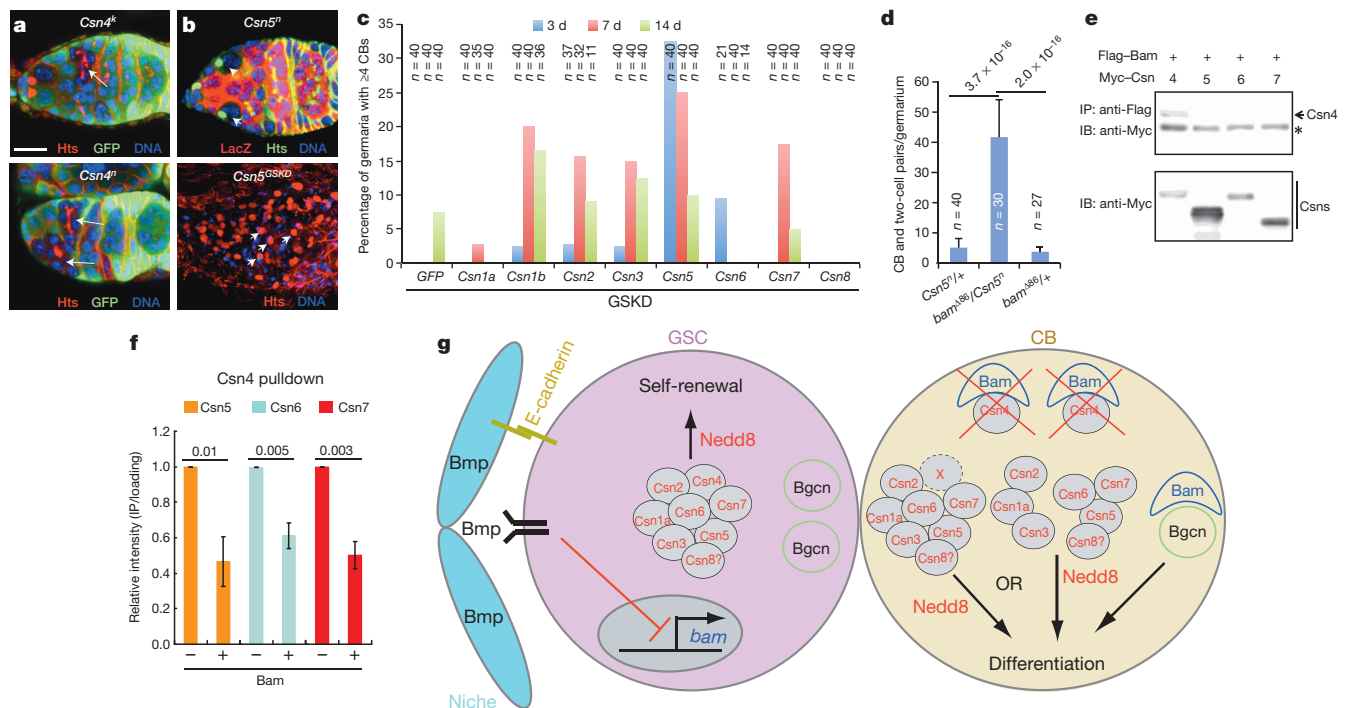


Figure 4 | Csn proteins promote cystoblast differentiation in the absence of Csn4. **a**, **b**, Marked *Csn4* mutant GSC progeny developed into 16-cell cysts (arrows, **a**), whereas marked *Csn5* mutant (b, top) or *Csn5* knock-down (b, bottom) GSC progeny remained as cystoblasts (arrowheads). Scale bar, 10 μ m. **c**, Germline-specific knock down of *Csn1b*, *Csn2*, *Csn3*, *Csn5* and *Csn7* significantly increased the number of germaria carrying four or more cystoblasts (CBs) in comparison with *GFP* and *Csn1a* knock-down negative controls. **d**, A heterozygous *Csn5^Δ* mutation significantly enhanced the

differentiation defect of the *bam^{Δ86}/+* mutant. **e**, Bam is associated only with Csn4, but not Csn5, when co-expressed in S2 cells, as determined by co-immunoprecipitation. The asterisk indicates a non-specific band. **f**, The presence of Bam protein significantly decreased the ability of Csn4 to co-immunoprecipitate with Csn5, Csn6 and Csn7 in S2 cells. **g**, A working model illustrating how Bam converts the function of the COP9 complex from self-renewal to differentiation by sequestering Csn4 via protein competition.

an excess of Bam, which neutralizes the self-renewal function of Csn4 and allows the formation of Bam-Bgcn protein complexes.

Although *Csn4* mutant GSC progeny develop normally into germ cell cysts, some of the *Csn5* mutant or knock-down GSC progeny failed to differentiate and accumulate as cystoblasts (Fig. 4a–c). In contrast to control germaria, which contained one or two cystoblasts, some of the *Csn1b*^{GSKD}, *Csn2*^{GSKD}, *Csn3*^{GSKD}, *Csn7*^{GSKD} and *Ned8*^{GSKD} germaria contained extra cystoblasts, exhibiting germ cell differentiation defects (Fig. 4c and Extended Data Fig. 8a–g). Consistent with this finding, the mutations in *Csn3* (*Csn3*^{FS}), *Csn5* (*csn5*ⁿ) and *Csn7* (*csn7*^c and *csn7*^{MB}) drastically and significantly enhanced the differentiation defect of the *bam* heterozygous mutant based on the number of cystoblasts and two-cell pairs (Fig. 4d and Extended Data Fig. 8h, i). Additionally, the *Csn4* and *bam* transheterozygous ovaries had significantly fewer cystoblasts and two-cell pairs than did single heterozygous ovaries, further suggesting that Csn4 can antagonize Bam function in differentiating GSC progeny (Extended Data Fig. 8k–m). We ruled out the possibility that the differentiation defects are caused by an excess of GSCs (Extended Data Fig. 8j, m'). Finally, mutations in *Csn5* and *Ned8* also drastically enhanced the differentiation defect of *bam*²/*bam*^{Δ86} (Extended Data Fig. 3c, d). Together, these results demonstrate that Csn proteins, except for Csn4, promote cystoblast differentiation.

One possible explanation for the opposite effects of Csn4 and other Csn proteins on cystoblast differentiation is that the sequestration of Csn4 by Bam allows other Csn proteins to have differentiation-promoting functions. The PCI (proteasome, COP9, initiation factor) domain in Csn4 is crucial for its assembly into COP9 through interaction with Csn5, Csn6 and Csn7 (refs 4, 25). Interestingly, Flag-Bam precipitated only Myc-Csn4 and not Myc-Csn5 and Myc-Csn6 in S2 cells, even in the presence of overexpressed Flag-Csn4, suggesting that Bam is associated only with Csn4 but not with the entire COP9 complex (Fig. 4e and Extended Data Fig. 9a). Additionally, the presence of Bam significantly interfered with the ability of Csn4 to interact with Csn5, Csn6 and Csn7 (Fig. 4f and Extended Data Fig. 9b–d). These results suggest that the sequestration of Csn4 by excess Bam in differentiating GSC progeny inactivates the self-renewal function of COP9, as well as allowing other Csn proteins to carry out differentiation-promoting functions.

This study has provided important insight into how protein competition controls the balance between GSC self-renewal and differentiation in the *D. melanogaster* ovary (Fig. 4g). In GSCs, Csn4 works within the COP9 complex to maintain GSC self-renewal. In differentiating GSC progeny, upregulated Bam proteins sequester Csn4 from the COP9 complex via protein competition, allowing other Csn proteins to promote germ cell differentiation, as well as inactivating COP9 self-renewal function. In addition, excess Bam can also form protein complexes with Bgcn to promote cystoblast differentiation by repressing self-renewal factors. Interestingly, Csn proteins have been identified to be required for maintaining human embryonic stem cells, as shown in a genome-wide RNA interference screen²⁶. Because many intrinsic self-renewal factors are expressed in both stem cells and in their differentiated progeny, protein competition could also be employed as a common mechanism for balancing stem cell self-renewal and differentiation in various stem cell systems.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 November 2013; accepted 6 June 2014.

Published online 6 August 2014.

1. Morrison, S. J. & Spradling, A. C. Stem cells and niches: mechanisms that promote stem cell maintenance throughout life. *Cell* **132**, 598–611 (2008).

2. Li, L. & Xie, T. Stem cell niche: structure and function. *Annu. Rev. Cell Dev. Biol.* **21**, 605–631 (2005).
3. Xie, T. Control of germline stem cell self-renewal and differentiation in the *Drosophila* ovary: concerted actions of niche signals and intrinsic factors. *Wiley Interdiscip. Rev. Dev. Biol.* **2**, 261–273 (2013).
4. Wei, N. et al. The COP9 complex is conserved between plants and mammals and is related to the 26S proteasome regulatory complex. *Curr. Biol.* **8**, 919–924 (1998).
5. Cope, G. A. & Deshaies, R. J. COP9 signalosome: a multifunctional regulator of SCF and other cullin-based ubiquitin ligases. *Cell* **114**, 663–671 (2003).
6. Lin, H. & Spradling, A. C. Germline stem cell division and egg chamber development in transplanted *Drosophila* germaria. *Dev. Biol.* **159**, 140–152 (1993).
7. Xie, T. & Spradling, A. C. Decapentaplegic is essential for the maintenance and division of germline stem cells in the *Drosophila* ovary. *Cell* **94**, 251–260 (1998).
8. Xie, T. & Spradling, A. C. A niche maintaining germ line stem cells in the *Drosophila* ovary. *Science* **290**, 328–330 (2000).
9. McKearin, D. & Ohlstein, B. A role for the *Drosophila* bag-of-marbles protein in the differentiation of cystoblasts from germline stem cells. *Development* **121**, 2937–2947 (1995).
10. Ohlstein, B. & McKearin, D. Ectopic expression of the *Drosophila* Bam protein eliminates oogenic germline stem cells. *Development* **124**, 3651–3662 (1997).
11. Shen, R., Weng, C., Yu, J. & Xie, T. eIF4A controls germline stem cell self-renewal by directly inhibiting BAM function in the *Drosophila* ovary. *Proc. Natl Acad. Sci. USA* **106**, 11623–11628 (2009).
12. Li, Y., Minor, N. T., Park, J. K., McKearin, D. M. & Maines, J. Z. Bam and Bgcn antagonize Nanos-dependent germ-line stem cell maintenance. *Proc. Natl Acad. Sci. USA* **106**, 9304–9309 (2009).
13. Chen, D. & McKearin, D. Dpp signaling silences *bam* transcription directly to establish asymmetric divisions of germline stem cells. *Curr. Biol.* **13**, 1786–1791 (2003).
14. Song, X. et al. Bmp signals from niche cells directly repress transcription of a differentiation-promoting gene, bag of marbles, in germline stem cells in the *Drosophila* ovary. *Development* **131**, 1353–1364 (2004).
15. Chen, D. & McKearin, D. M. A discrete transcriptional silencer in the *bam* gene determines asymmetric division of the *Drosophila* germline stem cell. *Development* **130**, 1159–1170 (2003).
16. Lin, H., Yue, L. & Spradling, A. C. The *Drosophila* fusome, a germline-specific organelle, contains membrane skeletal proteins and functions in cyst formation. *Development* **120**, 947–956 (1994).
17. Venken, K. J., He, Y., Hoskins, R. A. & Bellen, H. J. [acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in *D. melanogaster*. *Science* **314**, 1747–1751 (2006).
18. Oron, E. et al. COP9 signalosome subunits 4 and 5 regulate multiple pleiotropic pathways in *Drosophila melanogaster*. *Development* **129**, 4399–4409 (2002).
19. Song, X., Zhu, C. H., Doan, C. & Xie, T. Germline stem cells anchored by adherens junctions in the *Drosophila* ovary niches. *Science* **296**, 1855–1857 (2002).
20. Doronkin, S., Djagaeva, I. & Beckendorf, S. K. The COP9 signalosome promotes degradation of Cyclin E during early *Drosophila* oogenesis. *Dev. Cell* **4**, 699–710 (2003).
21. Ou, C. Y., Lin, Y. F., Chen, Y. J. & Chien, C. T. Distinct protein degradation mechanisms mediated by Cul1 and Cul3 controlling Ci stability in *Drosophila* eye development. *Genes Dev.* **16**, 2403–2414 (2002).
22. Djagaeva, I. & Doronkin, S. Dual regulation of dendritic morphogenesis in *Drosophila* by the COP9 signalosome. *PLoS ONE* **4**, e7577 (2009).
23. Schwechheimer, C. et al. Interactions of the COP9 signalosome with the E3 ubiquitin ligase SCFTIR1 in mediating auxin response. *Science* **292**, 1379–1382 (2001).
24. Ni, J. Q. et al. A genome-scale shRNA resource for transgenic RNAi in *Drosophila*. *Nature Methods* **8**, 405–407 (2011).
25. Hofmann, K. & Bucher, P. The PCI domain: a common theme in three multiprotein complexes. *Trends Biochem. Sci.* **23**, 204–205 (1998).
26. Chia, N. Y. et al. A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature* **468**, 316–320 (2010).

Acknowledgements We thank Y. Yamashita, the Developmental Studies Hybridoma Bank and the Bloomington *Drosophila* Stock Center for reagents, Xie laboratory members for discussions, and D. Chao and R. Krumlauf for critical comments on the manuscript. This work was supported by the Stowers Institute for Medical Research (T.X.), the National Institutes of Health (GM64428, T.X.), the National Natural Science Foundation of China (31370909, L.P.) and the Ministry of Science and Technology of China (2012CB518900, L.P.).

Author Contributions L.P., S.W., T.L. and T.X. conceived the project. L.P., S.W., T.L., C.W., X.S., J.K.P., J.Y., H.T., D.M.M. and T.X. collected and analysed the data. J.S., Z.-H.Y., D.A.C. and J.N. contributed the reagents. L.P., S.W., T.L. and T.X. prepared the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.X. (txg@stowers.org).

METHODS

D. melanogaster stocks. The following *D. melanogaster* stocks used in this study are described in FlyBase, unless specified: *bam*^{Z3-2884} (*bam*^Z), *bam*^{A86}, *bam*-GFP, *Csn4*^{K08018} (*Csn4*^K), *Csn4*^{EY08080} (*Csn4*^{EY}), *Csn7*^{MB01896} (*Csn7*^{MB}), *Csn7*^{E02176} (*Csn7*^E), *Csn3*^{FS}, *nos*-GAL4, UASp-*Csn4*, FRT42D, FRT82B, *armadillo*-lacZ, Ubi-GFP, *Csn4*^{null} (*Csn4*ⁿ), *Csn5*^{null} (*Csn5*ⁿ), *hs*-FLP, *nos*>STOP>GAL4 (provided by Y. Yamashita). The UASp-short hairpin RNA transgenic lines against GFP, *Csn1a*, *Csn2*, *Csn3*, *Csn5*, *Csn6*, *Csn7* and *Csn8* were inserted at the attP2 site on the third chromosome, as described previously²⁴. All stocks were cultured at 25 °C on standard cornmeal/molasses/agar medium unless specified.

Genetic screen for modifiers of the *bam* hypomorphic mutant combination. The sensitized *bam* genetic background *bam*^Z/*bam*^{A86} was used to test whether mutations in any of the 'stemness' genes identified in mammalian systems modify the differentiation defect of the *bam* mutant. In addition, mutations in the genes encoding various kinases and signalling molecules were also tested. In this screen, 11 enhancers were identified based on their ability to reduce the fecundity of *bam* mutant females, but *Csn4* was the only suppressor based on improvement in the low fecundity of *bam* mutant females.

Genetic clonal analysis. The marked control, *Csn4* and *Csn5* mutant GSC clones were generated using the FLP-mediated FRT recombination technique, as described previously⁷. The following genotypes were used for clonal analysis: (1) *hs*-flp/+; FRT42D/FRT42D Ubi-GFP; (2) *hs*-flp/+; FRT42D *Csn4*ⁿ/FRT42D Ubi-GFP; (3) *hs*-flp/+; FRT42D *Csn4*^K/FRT42D Ubi-GFP; (4) *hs*-flp/+; FRT82B/FRT82B *armadillo*-lacZ; (5) *hs*-flp/+; FRT82B *Csn5*ⁿ/FRT82B *armadillo*-lacZ; and (6) *hs*-flp/+; FRT42D *Csn4*ⁿ/FRT42D *armadillo*-lacZ; *bam*-GFP/+.

Germline-specific RNA-interference-mediated knock down (GSKD) of *Csn* genes. To knock down the functions of *Csn* genes in adult GSCs and their progeny, *hs*-flp; *nos*>STOP>GAL4 UAS-GFP/CyO virgins were crossed with males of UASp-RNAi lines against *Csn* genes and GFP. The female progeny were collected and heat-shocked in a 37 °C water bath (two 1-h treatments with an 8-h interval). Ovaries were isolated 3, 7 and 14 d AHS or 1, 7 and 13 d AHS for examining GSCs and cystoblasts after being labelled for Hts. For random sampling of flies, we anaesthetized the flies with carbon dioxide, mixed them and randomly picked enough females for ovary dissection. After the stained ovaries had been mounted on slides, we randomly selected and examined the germaria for GSCs and cystoblasts. All of the statistical analyses in this study were carried out by two-sided Student's *t*-test.

Generation of Flag-tagged *Csn4* and *Csn5* transgenes using the P[acman] system. To generate Flag-tagged *Csn4* and *Csn5* transgenes, the P[acman] BAC clones CH322-141C03 (*Csn4*) and CH322-113L01 (*Csn5*) were chosen because they contain the nearby genes on both the 5' and 3' sides and thus should contain all the regulatory sequences (Extended Data Fig. 2). The Flag-tagged versions of *Csn4* and *Csn5*, *Csn4*-Flag (C-terminal tag), *Flag*-*Csn5* (N-terminal tag) and *Csn5*-Flag (C-terminal tag), were generated by following previously reported experimental procedures²⁷. Confirmed plasmids were further used to make transgenic fly strains.

BrdU labelling of GSCs. At 1 week ACI, the freshly isolated *D. melanogaster* ovaries were incubated with 75 µg ml⁻¹ BrdU in Grace's Insect Medium for 1 h at room temperature. Then, the ovaries were fixed and processed for immunostaining with anti-BrdU and anti-LacZ or anti-GFP antibodies for the detection of clones according to our published procedures⁷.

Immunohistochemistry. The following antisera were used: monoclonal mouse anti-Hts antibody 1B1 (1:3; DSHB); rabbit polyclonal anti-β-galactosidase antibody (1:300; Cappel); monoclonal mouse anti-β-galactosidase antibody (1:100; Promega); rabbit polyclonal anti-p-Smad3 (1:200; Epitomics); mouse monoclonal anti-BrdU antibody (1:20; Sigma-Aldrich); rat monoclonal anti-E-cadherin antibody (1:4; DSHB); rabbit polyclonal anti-GFP antibody (1:200; Molecular Probes); and rat anti-Vasa

antibody (1:3; DSHB). Secondary antibodies including goat anti-rabbit, anti-mouse or anti-rat IgG conjugated to Alexa Fluor 488 or 568 (Molecular Probes) were used at 1:200. All micrographs were taken with Leica confocal microscopes.

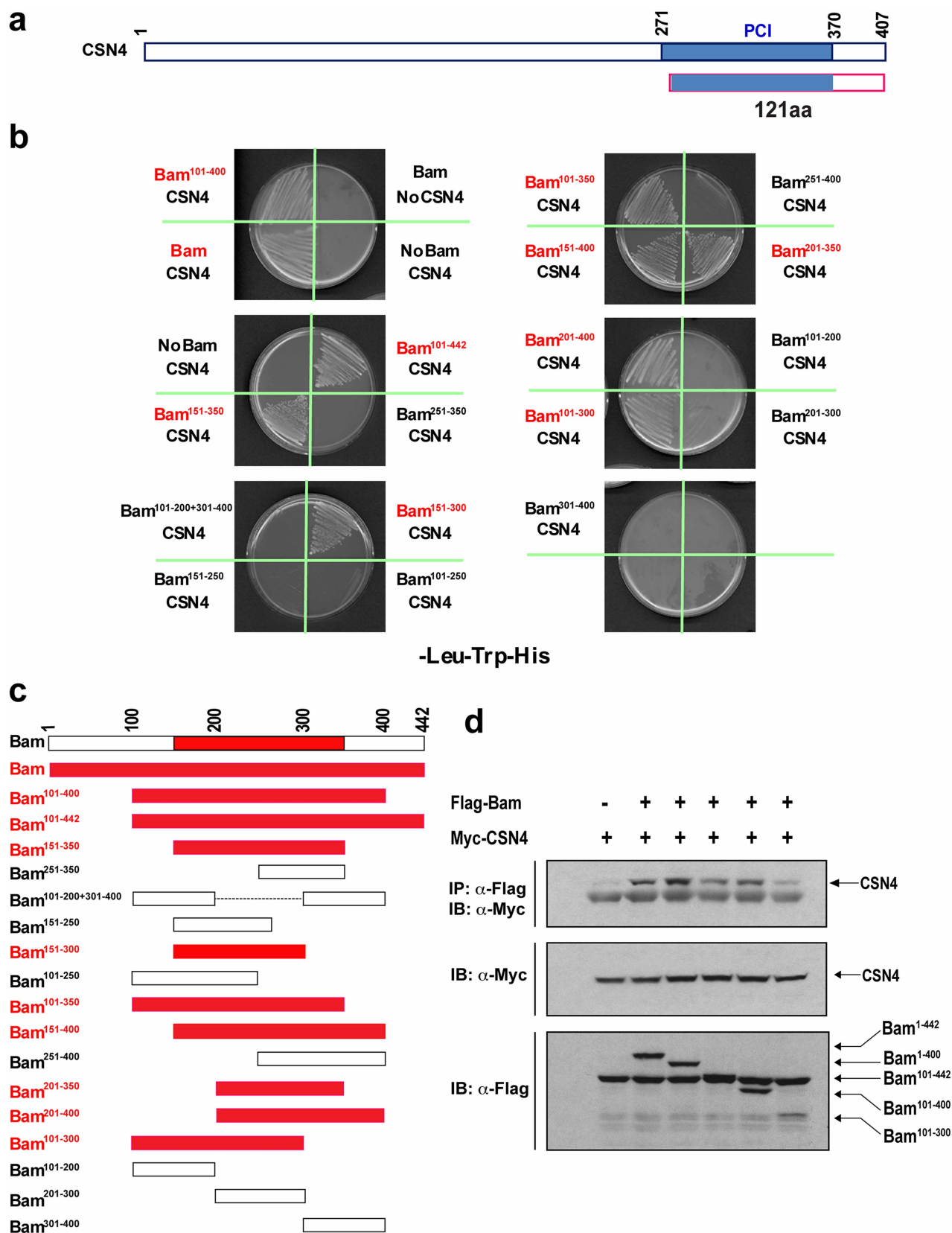
Plasmid construction. Invitrogen Gateway technology was used to construct the Myc- and HA-tagged *Csn4*, Myc-tagged *Csn5*, Myc-tagged Bgcn and Flag-tagged Bam plasmids for expression in S2 cells and for making transgenic flies. The coding regions or various truncations of *bam*, *bgcn*, *Csn4* and *Csn5*, which were amplified from a *D. melanogaster* ovarian cDNA library, were cloned into the pENTR TOPO cloning vector and completely sequenced. These pENTR vectors were subsequently recombined with Flag-, Myc- or HA-tagged destination vectors (from T. Murphy) by using LR Clonase (Invitrogen).

Co-immunoprecipitation and western blotting. *D. melanogaster* S2 cells (Invitrogen) were cultured at 25 °C in Schneider's *Drosophila* Medium according to the manufacturer's manual. S2 cells were transfected with the indicated amounts of plasmids using Cellfectin (Invitrogen) according to the manufacturer's manual. The S2 cells were harvested 36–48 h after transfection for immunoprecipitation and western blot analysis. For immunoprecipitation, 4 × 10⁶ cells were lysed with 800 µl ice-cold lysis buffer (20 mM Tris.HCl, pH 7.5, 100 mM NaCl, 1% Nonidet P-40, 1 mM EDTA, 10% glycerol and a mixture of protease inhibitors). For immunoprecipitation experiments, anti-Flag monoclonal antibody M2 agarose (Sigma) was preincubated with 5% BSA at 4 °C for 1 h, and then 50 µl 50% agarose slurry was used in each experiment. For immunoprecipitation with monoclonal anti-Myc and anti-HA antibodies, cell lysates were incubated with 2 µg anti-Myc antibody (C29F4, Cell Signaling Technology) at 4 °C overnight, and then 20 µl precleared Protein A+G agarose (Sigma) was added in each experiment. Then, the immunoprecipitation was carried out exactly as previously described²⁸. Monoclonal anti-Myc (9E10, Sigma) and monoclonal anti-Flag M2 peroxidase (Sigma) antibodies were used for western blots.

Yeast two-hybrid screen and interaction assays. The coding region and various truncations of *bam* amplified from a *D. melanogaster* ovarian cDNA library were cloned into the pGBKT7 vector as the baits and completely sequenced. The coding regions of *bgcn* and *Csn4*, which were amplified from a *D. melanogaster* ovarian cDNA library, were cloned into the pGADT7 vector and completely sequenced. The baits pGBKT7-*bam*/truncations and pGADT7-*bgcn* or pGADT7-*Csn4* were cotransformed into the Y187 yeast strain to make the two-hybrid system. Positive interactions were screened using Leu-Trp-His triple selection markers.

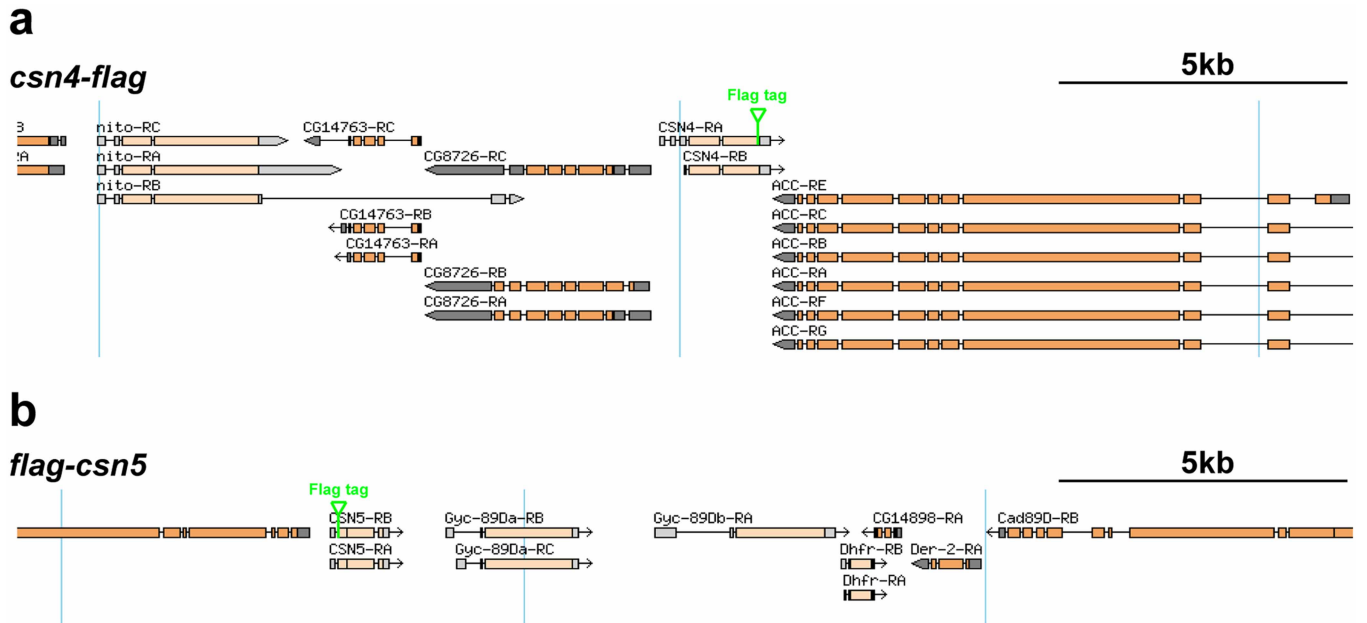
For α-galactosidase unit measurement, interaction samples (Bam-Csn4 and Bam-Bgcn) were cultured in synthetic defined medium without His, Leu and Trp, whereas negative controls (BK-Csn4, BK-Bgcn and Bam-AD) were cultured in synthetic defined medium without Leu and Trp. Supernatant from the liquid culture was used to quantify yeast extracellular α-galactosidase activity resulting from expression of the *MEL1* reporter gene in strain AH109. The catalytic activity of α-galactosidase was monitored colorimetrically by measuring the rate of hydrolysis of the chromogenic substrate PNP-α-Gal (N0877, Sigma) at 405 nm. This assay is a sensitive colorimetric method for quantifying the interaction in a two-hybrid system using strain AH109. α-galactosidase = (mU ml⁻¹ × cell number) = OD₄₀₅ × V_f × 1,000/(10.5t × V_i × OD₆₀₀), where OD is optical density, V_f is the final volume of the assay, V_i is the volume of culture medium supernatant added and *t* is the elapsed incubation time.

27. Venken, K. J. T. et al. Versatile P[acman] BAC libraries for transgenesis studies in *Drosophila melanogaster*. *Nature Methods* **6**, 431–434 (2009).
28. Yu, J. et al. MTMR4 attenuates transforming growth factor β (TGFβ) signaling by dephosphorylating R-Smads in endosomes. *J. Biol. Chem.* **285**, 8454–8462 (2010).



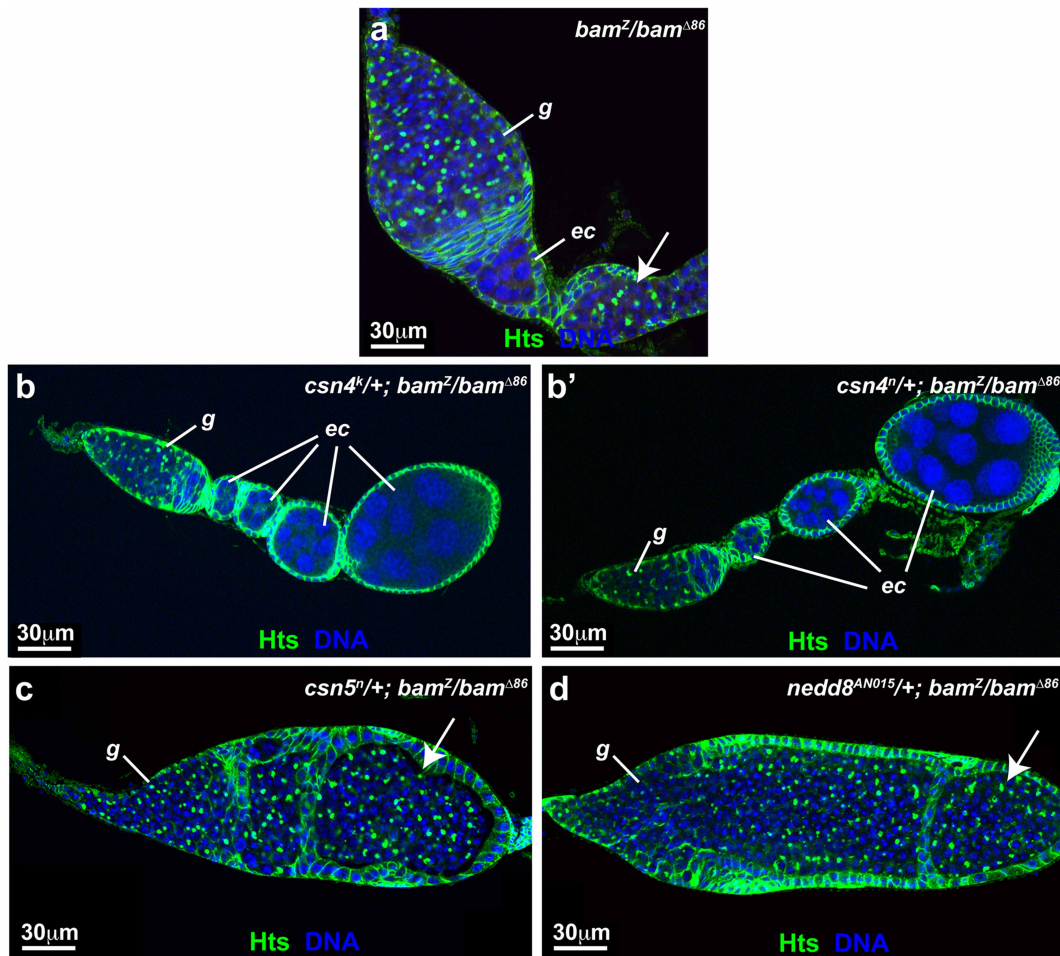
Extended Data Figure 1 | Bam interacts with Csn4 in yeast cells. **a**, Csn4 contains a PCI/PINT protein interaction domain in its C terminus. The C-terminal 121 amino acid region containing the PCI/PINT domain interacts with Bam in yeast cells. **b**, **c**, In a yeast two-hybrid assay, different Bam truncations fused to the Gal4 DNA-binding domain were individually tested for their interaction with the fusion protein between Csn4 and the Gal4 transcriptional activation domain by activating expression of the *HIS3* gene,

which allows yeast cells to grow on medium lacking histidine. “No Bam” means that only the Gal4 binding domain is used in the assay. The results from **b** are summarized in **c**. Red-filled boxes indicate positive interactions, whereas unfilled boxes denote no interaction. **d**, Co-immunoprecipitation (co-IP) experiments show that Myc-tagged Csn4 interacted with different truncated Bam proteins tagged by Flag in S2 cells.



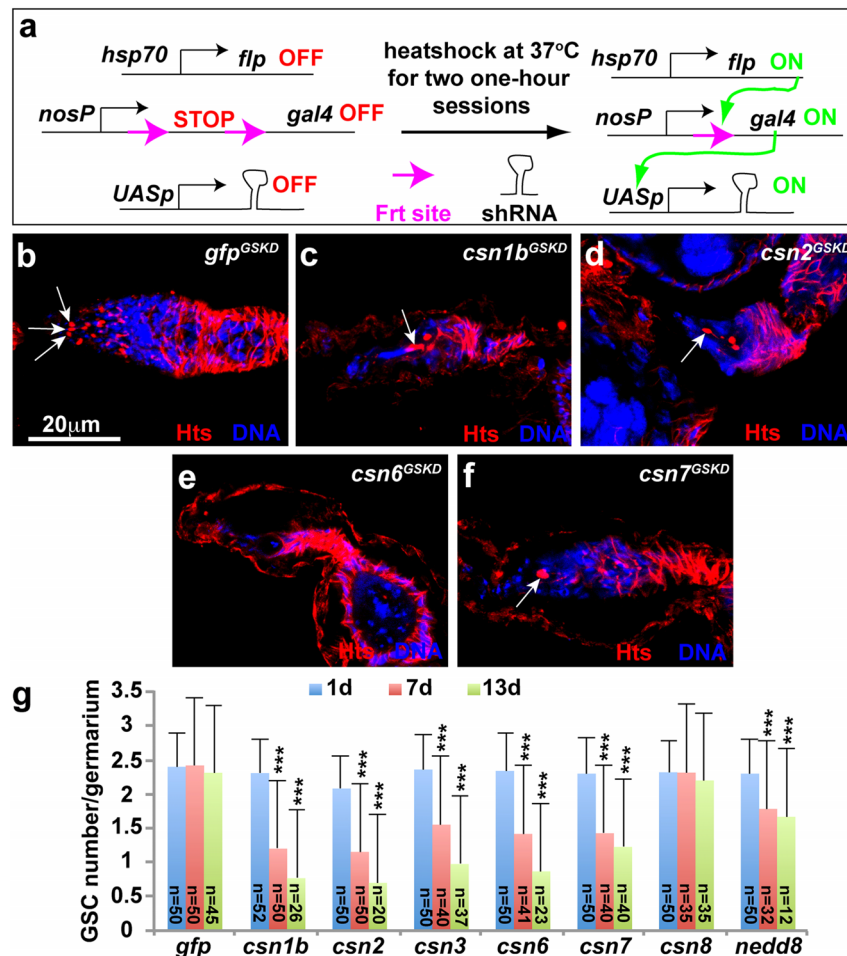
Extended Data Figure 2 | BAC transgenes for Flag-tagged *Csn4* and *Csn5*.
a, The BAC clone contains the 21-kilobase genomic region in which *Csn4* is flanked by three genes at the 5' end and one gene at the 3' end. A Flag-tag sequence (green symbol) is inserted before the stop codon (*Csn4-Flag*). **b**, The

BAC clone contains the 21-kilobase genomic region in which *Csn5* is flanked by one gene at the 5' end and four genes at the 3' end. A Flag-tag sequence (green symbol) is inserted after the translation initiation codon, ATG (*Flag-Csn5*).



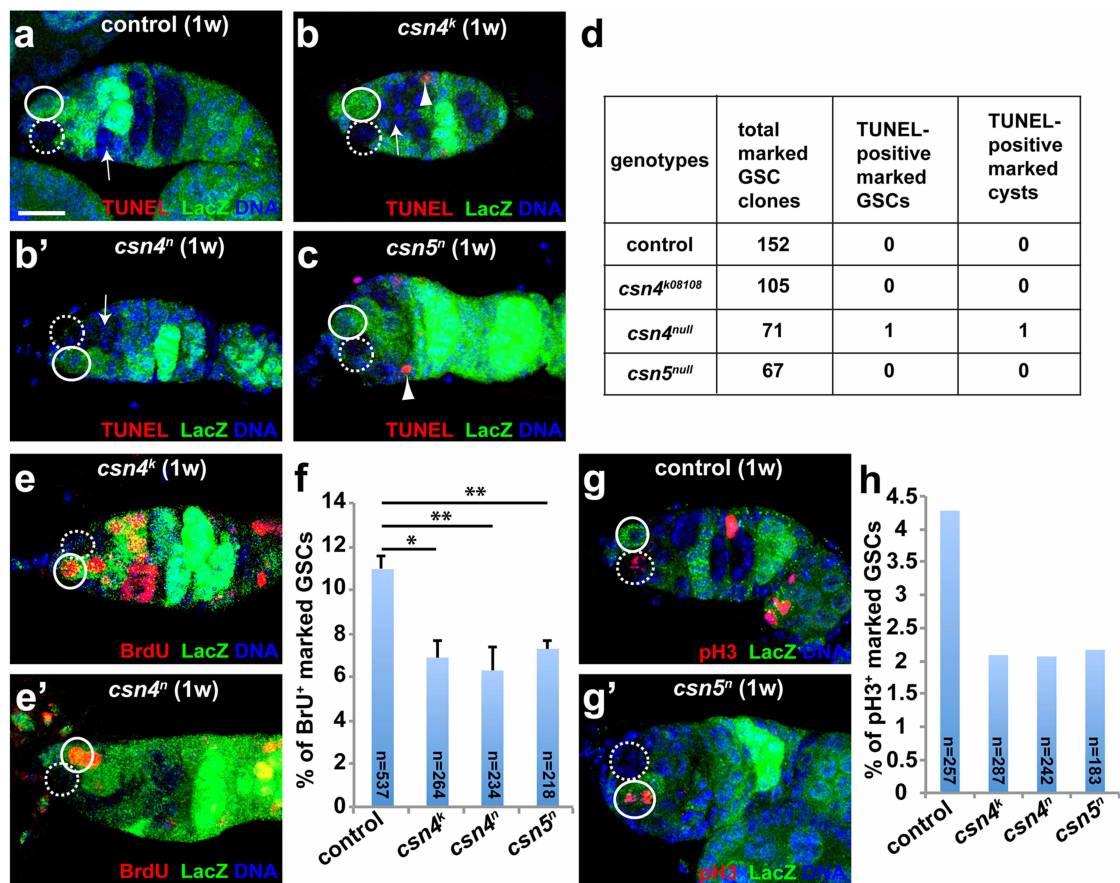
Extended Data Figure 3 | Mutations in *Csn4* suppress, but mutations in *Csn5* and *Nedd8* enhance, the *bam* differentiation defect. The germarium is indicated by “g”. **a**, A *bam^Z/bam^{Δ86}* mutant ovariole contains a tumorous germarium followed by a normal egg chamber (ec) and a tumorous egg chamber (arrow) (the same as in Fig. 1e). **b**, **b'** *bam^Z/bam^{Δ86}* mutant ovarioles

that are also heterozygous for *Csn4* contain tumorous germaria and three or more normal egg chambers. **c**, **d**, *bam^Z/bam^{Δ86}* mutant ovarioles that are also heterozygous for *Csn5* (**c**) or *Nedd8* (**d**) contain tumorous germaria and tumorous egg chambers (arrows).



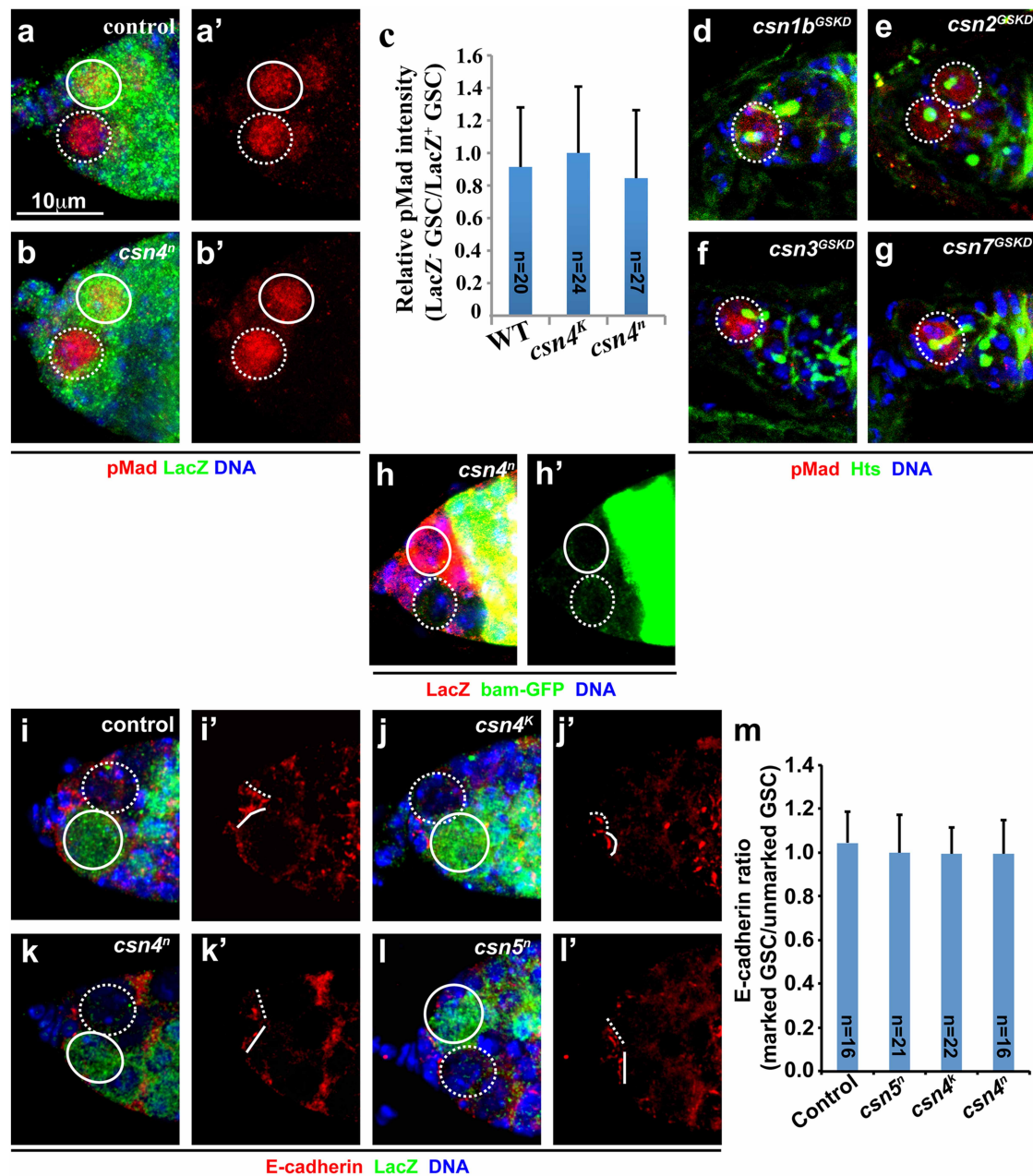
Extended Data Figure 4 | Germline-specific knock down of *Csn1b*, *Csn2*, *Csn6*, *Csn7* and *Csn8* in the adult ovary leads to GSC loss. **a**, The combined FLP-Out and Gal4-UAS system for knocking down gene function in adult GSCs and their progeny. An FRT-STOP (transcription stop sequence)-FRT cassette is inserted between the germ-cell-specific *nos* promoter and the *GAL4* coding region to prevent Gal4 expression in germ cells. With heat-shock treatments, FLP expression removes the STOP cassette in most GSCs, leading

to Gal4 expression and then short hairpin RNA expression. **b-f**, Germ-cell-specific *Csn1b* (**c**), *Csn2* (**d**), *Csn6* (**e**) and *Csn7* (**f**) germaria contained one (arrow in **c**, **d** and **f**) or no (**e**) GSCs in contrast to the *GFP* knock-down germarium, which carried three GSCs (**b**, arrows). **g**, The results from the independent knock-down experiment show that germline-specific knock down of *Csn1b*, *Csn2*, *Csn3*, *Csn6*, *Csn7* and *Nedd8* caused significant GSC loss in comparison with the *GFP* knock-down negative control. ***, $P < 0.001$.



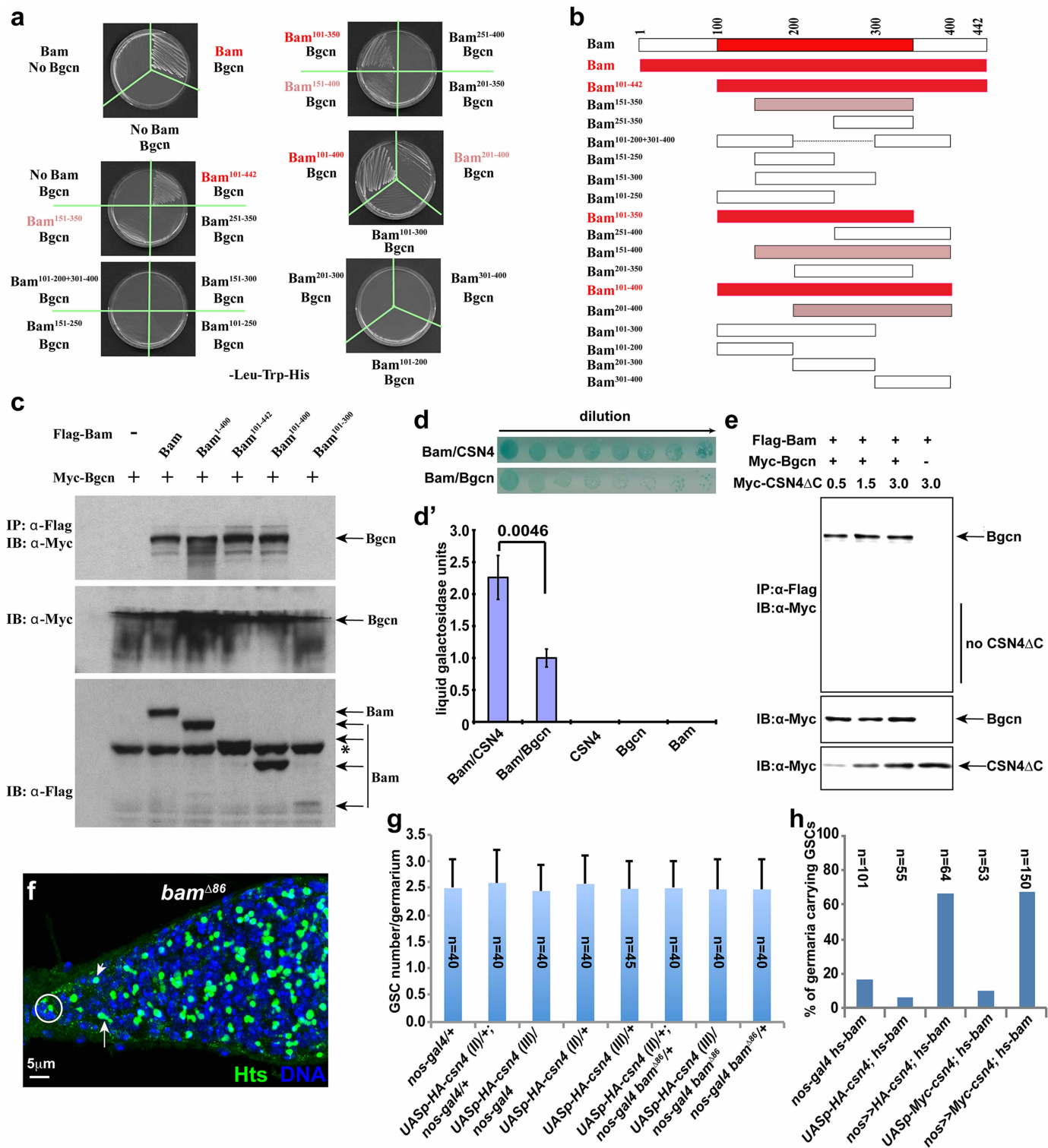
Extended Data Figure 5 | Marked *Csn4* and *Csn5* mutant GSCs are not prone to apoptosis but proliferate more slowly than marked control GSCs. Marked GSCs and cysts are indicated by dashed circles and arrows, respectively. Unmarked GSCs are highlighted by circles. **a**, The marked control GSC and cysts are negative for TUNEL labelling. **b**, **b'**, The marked *Csn4^k* (**b**) and *Csn4ⁿ* (**b'**) mutant GSCs and cysts are negative for TUNEL labelling. **c**, The marked *Csn5ⁿ* mutant GSCs are negative for TUNEL labelling. Arrowheads indicate

dying somatic cells. **d**, Quantification of TUNEL-positive marked GSCs and cysts. **e**, **f**, Marked mutant *Csn4^k* or *Csn4ⁿ* mutant GSCs show significantly lower BrdU incorporation rates than marked control GSCs. Quantification is shown in **f** (**, $P < 0.01$; *, $P < 0.05$). **g**, **h**, In contrast to the marked control GSCs (**g**), the marked *Csn4* and *Csn5* mutant GSCs have low p-H3 labelling rates (**g'**). Quantification is shown in **h**.

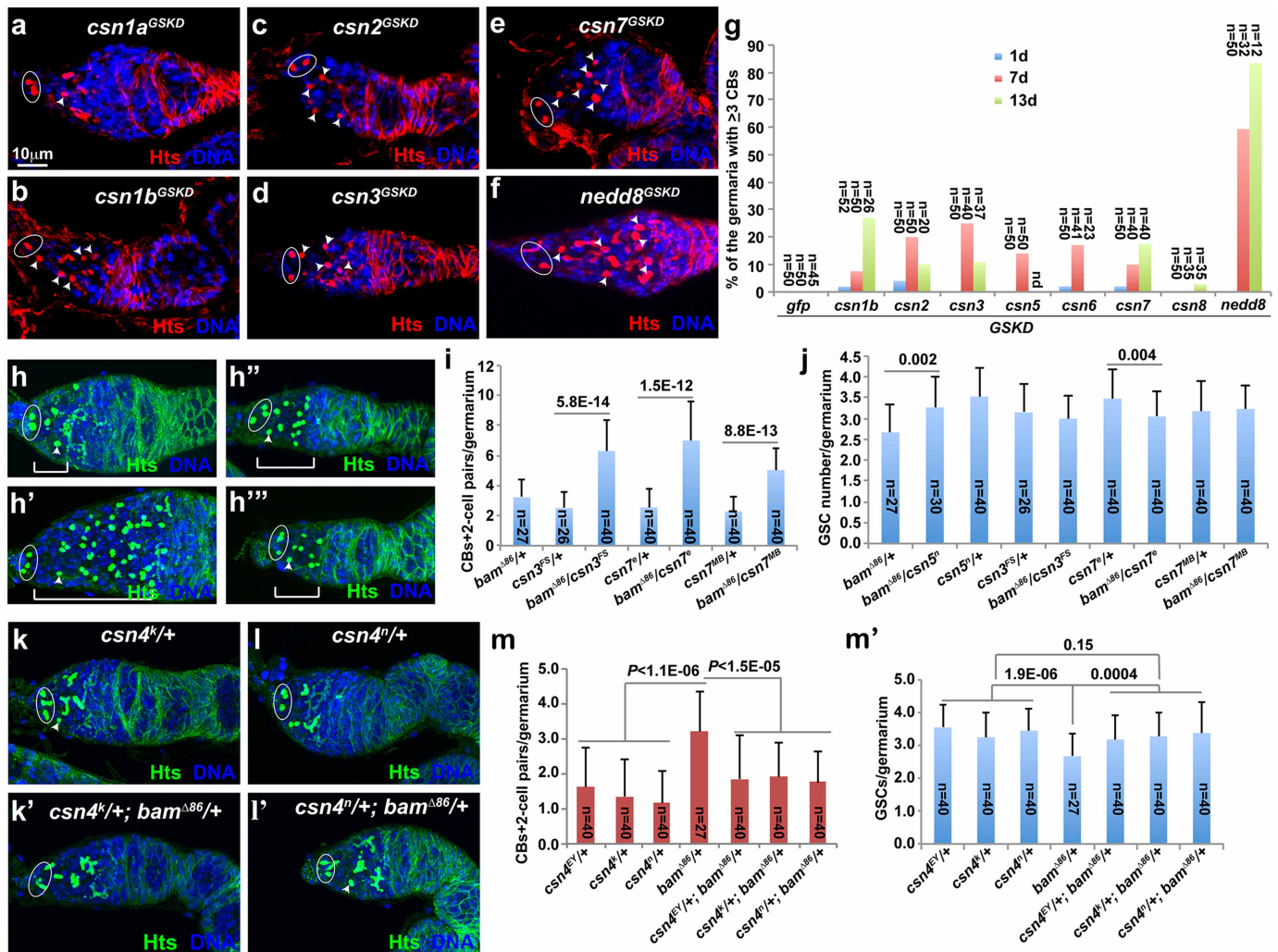


Extended Data Figure 6 | Inactivation of Csn functions in GSCs does not affect Bmp signalling and E-cadherin expression. **a, a'**, Marked (dashed circle) and unmarked (solid circle) control GSCs express p-Mad at similar levels. **b, b'**, Marked *Csn4ⁿ* mutant (dashed circle) and unmarked control (solid circle) GSCs express p-Mad at similar levels. **c**, Quantification results show that marked *Csn4^K* and *Csn4ⁿ* mutant GSCs express similar p-Mad levels to marked control GSCs. **d–g**, p-Mad is expressed normally in *Csn1b^{GSKD}*

(**d**), *Csn2^{GSKD}* (**e**), *Csn3^{GSKD}* (**f**) and *Csn7^{GSKD}* (**g**) GSCs (dashed circles). **h, h'**, *bam-GFP* is repressed in LacZ-negative *Csn4ⁿ* mutant (dashed circle) and LacZ-positive unmarked control (solid circle) GSCs. **i–m**, LacZ-negative *Csn4^K* (**j**), *Csn4ⁿ* (**k**) and *Csn5ⁿ* (**l**) mutant GSCs (dashed circles) express similar levels of E-cadherin in the stem-cell-niche junction to unmarked control GSCs (solid circles, **j–l**), as well as marked and unmarked control GSCs (dashed and solid circles, **i**). Quantification results are shown in **m**.

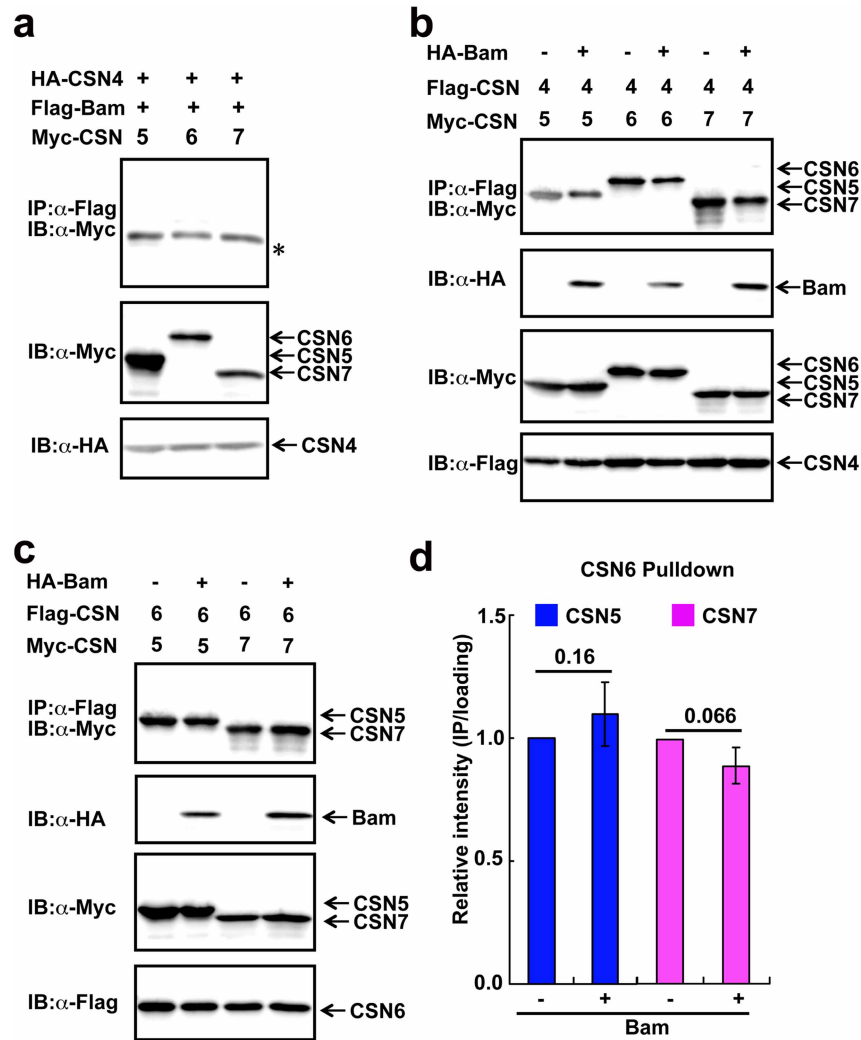


interacted significantly more strongly with Csn4 than with Bgcn. **e**, The C-terminal deletion of the Bam-interacting domain (the C-terminal 121 amino acids) of Csn4 protein destroys Csn4's ability to compete with Bgcn for binding to Bam. **f**, A *bam*^{Δ86} homozygous germline contains two GSCs (circle); single germ cells are identified by the spherical spectrosome (arrowhead) and two-cell pairs by the elongated fusome (arrow). These two-cell pairs undergo complete cytokinesis to generate two cystoblasts. **g**, GSC quantification results for the experiments in Fig. 3d show that germline-specific expression of Csn4 did not change the GSC number in either wild-type or *bam* heterozygous ovaries. **h**, Quantification results for Fig. 3e show that Csn4 overexpression suppressed ectopic Bam-expression-induced GSC loss.



Extended Data Figure 8 | Csn proteins other than Csn4 promote cystoblast differentiation. **a–f**, *Csn1b^{GSKD}* (**a**), *Csn2^{GSKD}* (**b**), *Csn3^{GSKD}* (**c**), *Csn7^{GSKD}* (**d**) and *Nedd8^{GSKD}* (**e**) and *Nedd8^{GSKD}* (**f**) germlaria contained four or more cystoblasts (arrowheads) in contrast to the one cystoblast in the *Csn1a^{GSKD}* germlarium (**a**). The ovals highlight GSCs. **g**, Quantification results from the independent knock-down experiment (from Fig. 4c) show that germline-specific knock down of *Csn1b*, *Csn2*, *Csn3*, *Csn5*, *Csn6*, *Csn7* and *Nedd8* significantly increased the number of germlaria carrying three or more cystoblasts in comparison with the *GFP* knock-down negative control. *Nedd8* knock down yielded the most severe differentiation defect. **h–i**, The *bam^{Δ86}/Csn5ⁿ* (**h'**), *bam^{Δ86}/Csn3^{FS}* (**h''**) and *bam^{Δ86}/csn7^{MB}* (**h'''**) transheterozygous germlaria had significantly more cystoblasts (arrowhead) plus two-cell pairs in their anterior region

(bracket) than did the *Csn5ⁿ* (**h**) heterozygous germlarium. Quantification results are shown in **i**, **j**, GSC quantification results show that *Csn* and *bam* transheterozygous ovaries did not contain significantly more GSCs than heterozygous *Csn* mutant ovaries. **k–l'**, *Csn4^k/+* (**k**) and *Csn4ⁿ/+* (**l**) heterozygous germlaria, as well as *csn4^k/+; bam^{Δ86}/+* (**k'**) and *csn4ⁿ/+; bam^{Δ86}/+* (**l'**) transheterozygous germlaria contained one cystoblast (arrowhead) or one two-cell pair (arrow). Ovals highlight GSCs. **m**, **m'**, Quantification results show that mutations in *Csn4*, *Csn4^{EY}*, *Csn4^k* and *Csn4ⁿ* suppressed the differentiation defect of the *bam* heterozygous mutant, despite *Csn4* and *bam* transheterozygotes having more GSCs than *bam* heterozygotes.



Extended Data Figure 9 | Bam interferes with the interaction between CSN4 and other Csn proteins. **a**, Bam did not associate with Csn5, Csn6 and Csn7 when co-expressed in S2 cells in the presence of Csn4. The asterisk indicates a non-specific band. **b**, The presence of the Bam protein decreased the

ability of Csn4 to co-IP with Csn5, Csn6 and Csn7 in S2 cells. **c**, The presence of the Bam protein did not affect the ability of Csn6 to co-IP with Csn5 and Csn7. **d**, Quantification results for three independent experiments shown in **c**.

Interleukin-22 alleviates metabolic disorders and restores mucosal immunity in diabetes

Xiaoting Wang^{1*}, Naruhisa Ota^{1*}, Paolo Manzanillo¹, Lance Kates², Jose Zavala-Solorio², Celine Eidenschenk¹, Juan Zhang¹, Justin Lesch¹, Wyne P. Lee¹, Jed Ross², Lauri Diehl³, Nicholas van Bruggen², Ganesh Kolumam^{2*} & Wenjun Ouyang¹

The connection between an altered gut microbiota and metabolic disorders such as obesity, diabetes, and cardiovascular disease is well established^{1,2}. Defects in preserving the integrity of the mucosal barriers can result in systemic endotoxaemia that contributes to chronic low-grade inflammation, which further promotes the development of metabolic syndrome^{3–5}. Interleukin (IL)-22 exerts essential roles in eliciting antimicrobial immunity and maintaining mucosal barrier integrity within the intestine^{6,7}. Here we investigate the connection between IL-22 and metabolic disorders. We find that the induction of IL-22 from innate lymphoid cells and CD4⁺ T cells is impaired in obese mice under various immune challenges, especially in the colon during infection with *Citrobacter rodentium*. While innate lymphoid cell populations are largely intact in obese mice, the upregulation of IL-23, a cytokine upstream of IL-22, is compromised during the infection. Consequently, these mice are susceptible to *C. rodentium* infection, and both exogenous IL-22 and IL-23 are able to restore the mucosal host defence. Importantly, we further unveil unexpected functions of IL-22 in regulating metabolism. Mice deficient in IL-22 receptor and fed with high-fat diet are prone to developing metabolic disorders. Strikingly, administration of exogenous IL-22 in genetically obese leptin-receptor-deficient (*db/db*) mice and mice fed with high-fat diet reverses many of the metabolic symptoms, including hyperglycaemia and insulin resistance. IL-22 shows diverse metabolic benefits, as it improves insulin sensitivity, preserves gut mucosal barrier and endocrine functions, decreases endotoxaemia and chronic inflammation, and regulates lipid metabolism in liver and adipose tissues. In summary, we identify the IL-22 pathway as a novel target for therapeutic intervention in metabolic diseases.

IL-22 is an IL-10 family cytokine expressed predominantly by innate lymphoid cells (ILCs) and CD4⁺ T helper (TH) subtypes such as TH17 and TH22 cells^{6,7}. IL-22 elicits innate host defence, and tissue protective and regenerative functions from various epithelial cells including hepatocytes and acinar cells. Mice deficient in IL-22 are compromised not only in controlling pathogenic infection, but also in maintaining intestinal epithelial integrity and homeostasis of commensals, which may result in systemic dissemination of bacteria leading to chronic inflammation^{6,9–11}. Interestingly, obese mice are commonly associated with chronic inflammation and deficiency in mucosal host defence^{1,12}, while mice with deficient mucosal immunity such as *Thr5*^{−/−} mice develop metabolic disorder⁵. Because of the reciprocal relationship between metabolism and mucosal immunity, we investigated the role of IL-22 in obesity-associated disorder and mucosal immunity. We began by examining whether the production of IL-22 was altered in mouse models of obesity. Although IL-22 expression from ILCs and CD4⁺ T cells under homeostatic conditions was largely normal (data not shown), we observed reduced IL-22 production from ILCs upon *in vivo* administration with flagellin¹³, and from CD4⁺ T cells after immunization with ovalbumin in genetically obese leptin-receptor (*db/db*)-, leptin (*ob/ob*)-deficient

mice, and in diet-induced obese (DIO) C57BL/6 mice fed with high-fat diet (HFD) (Extended Data Fig. 1a–i). Leptin itself did not directly enhance IL-22 production from either ILCs or T cells *in vitro* (data not shown). Previous studies demonstrated that immune responses such as the generation of TH subsets under specific immune challenges were compromised in leptin-deficient mice^{14–17}. However, under our experimental conditions, only IL-17⁺ TH17 cells besides IL-22⁺ T cells, but not other TH subsets, were significantly decreased (Extended Data Fig. 1j–m). Importantly, upon infection with the intestinal pathogen *C. rodentium*, a dramatic reduction of peak IL-22 expression was observed in the colon of DIO, *ob/ob*, and, to a lesser extent, *db/db* mice (Fig. 1a, b). Taken together, these data suggest a defect in IL-22 induction during infection and immune challenges in obese mice.

IL-22 is indispensable for mucosal defence against *C. rodentium* infection⁹. To determine further the functional consequences of reduced IL-22 signalling in obese mice, we analysed the survivability and disease progression of obese mice infected with *C. rodentium*. Although there was no significant weight loss in *db/db* mice after oral inoculation with *C. rodentium* (Fig. 1c), 60–100% of *db/db* mice succumbed during the second week of the infection in repeated experiments (Fig. 1d and data not shown). Histological analysis of the colon sections from infected *db/db* mice revealed increased leukocyte infiltration, severe epithelial damage, and shedding at the mucosal surface (Fig. 1e–i). In addition, these mice showed patchy submucosal oedema and multifocal bacterial colonies, which were often associated with localized necrosis. Significantly elevated bacterial burdens were also detected both in the liver and spleen of *db/db* mice (Fig. 1j, k). Similar results were also observed in *ob/ob* mice (Extended Data Fig. 2a–g). Interestingly, despite the reduction of IL-22 expression, DIO mice survived *C. rodentium* infection (Extended Data Fig. 2h), probably because of the older age of these mice (~18 weeks) at the time of the infection compared with that of leptin-pathway-deficient mice (5–6 weeks). However, while *C. rodentium* was already cleared in the colon of chow-diet-fed mice by 18 days after infection, a high bacterial burden was still detected in infected DIO mice, suggesting a similar mucosal defence defect in these mice (Extended Data Fig. 2i). Lastly, the antibody production against *C. rodentium*, which is required for the clearance of the bacteria from the host during the later phase of the infection^{14,18}, was compromised in *db/db* mice (Fig. 1l). Hence, the failed host defence against *C. rodentium* in obese mice was probably caused by defects both in the production of IL-22 from ILCs and the adaptive antibody response.

During *C. rodentium* infection, IL-22 is primarily produced by ILCs upon activation by the cytokine IL-23 (refs 7, 9). To understand the mechanism of reduced IL-22 expression in obese mice, we first examined whether ILCs and T cells had an intrinsic inability to produce IL-22 upon stimulation. In *db/db* and *ob/ob* mice, the numbers of IL-22⁺ ILCs and T cells isolated from infected colons were comparable to those in wild-type (WT) mice upon *in vitro* re-stimulation, suggesting that the ability to induce IL-22 remained intact (Extended Data Fig. 3a).

¹Department of Immunology, Genentech, South San Francisco, California 94080, USA. ²Department of Biomedical Imaging, Genentech, South San Francisco, California 94080, USA. ³Department of Pathology, Genentech, South San Francisco, California 94080, USA.

*These authors contributed equally to this work.

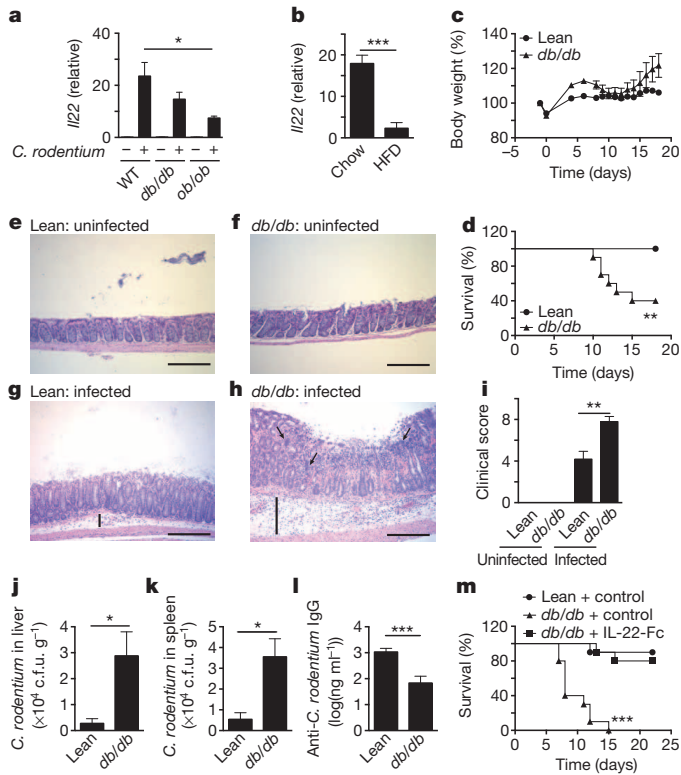


Figure 1 | Susceptibility of obese mice to *C. rodentium* infection is associated with defective IL-22 induction. **a, b**, IL22 messenger RNA (mRNA) expression in colons from WT, *db/db*, and *ob/ob* mice ($n = 5$) (**a**) and chow-diet-fed and HFD-fed mice ($n = 7$) (**b**) at 4 days after *C. rodentium* infection. **c, d**, Body weight (**c**) and survival (**d**) of *db/db* and lean control mice infected with *C. rodentium* ($n = 10$). **e–h**, Colon histology of lean control (**e, g**) and *db/db* (**f, h**) mice before (**e, f**) or 10 days after (**g, h**) infection. Arrows, bacterial colonies; vertical bar, submucosal oedema; horizontal bar, 200 μm . **i**, Clinical score determined by colon histology ($n = 5$). **j, k**, Bacterial burden in liver (**j**) and spleen (**k**) on day 10 ($n = 5$). **l**, Anti-*C. rodentium* immunoglobulin G (IgG) in serum ($n = 5$) on day 10. **m**, Survival of lean control and *db/db* mice infected and treated with IL-22-Fc or control IgG ($n = 10$). Error bars, s.e.m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, unpaired Student's *t*-test (**a, b, i–l**) or log rank test (**d, m**). Data shown are representative of three independent experiments.

Interestingly, we noticed a reduced expression of IL-23 p19 and IL-12/23 p40 subunits in the colon of leptin-pathway-deficient mice 4 days after infection with *C. rodentium* (Extended Data Fig. 3b, c). Furthermore, systemic overexpression of IL-23 by hydrodynamic tail vein injection restored the expression of IL-22 in the colon and prevented mortality in *ob/ob* mice (Extended Data Fig. 3d, e), supporting the premise that the defect in IL-23 induction contributed directly to the reduced production of IL-22 and morbidity upon infection. Although dendritic cell subsets were identified as primary sources of IL-23, the exact pathways required for IL-23 induction during *C. rodentium* infection in wild type remained unclear. Further research is necessary to elucidate the defective IL-23 production in obese mice. Finally, administration of exogenous IL-22-Fc was able to reduce mortality (Fig. 1m), alleviate epithelial damage and inflammation, and inhibit dissemination of bacteria in liver and spleen in *db/db* mice infected with *C. rodentium* (Extended Data Fig. 3f–h). Together, these data support the essential roles of the IL-22 pathway in mucosal defence in obese mice.

Given the essential connection between mucosal immunity and obesity¹, we examined whether lack of IL-22 signalling contributed to the development of metabolic syndrome. We first examined the development of metabolic syndrome both in IL-22- and IL-22R1-deficient mice (Extended Data Fig. 4a–c) fed with HFD. Relative to control mice, IL-22R1-deficient mice displayed higher weight gain after 1 month of HFD and

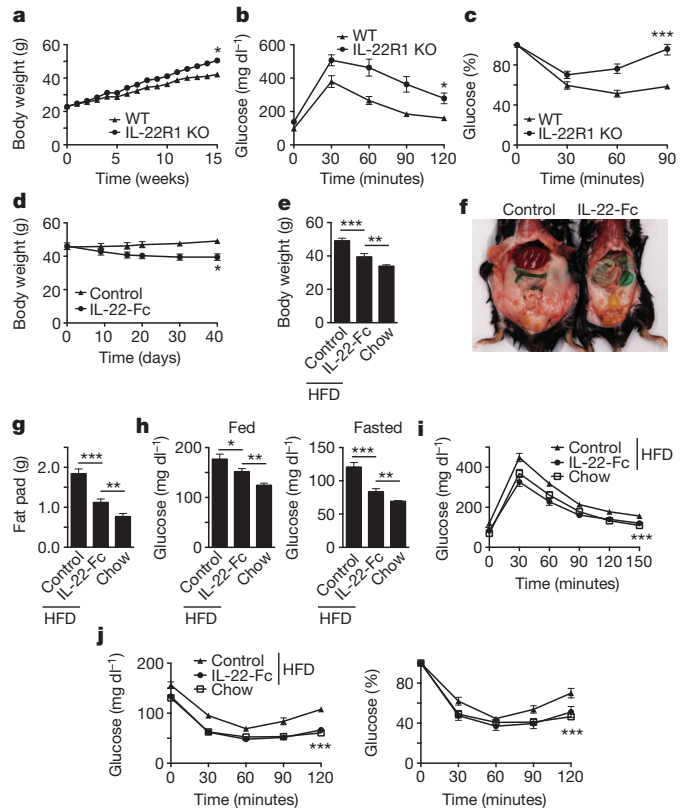


Figure 2 | The IL-22 pathway is essential in controlling metabolic disorders. **a–c**, IL-22R1 knockout (KO) ($n = 9$) and WT male littermates ($n = 6$) were fed with HFD. **a**, Body weight. **b**, GTT at 3 months with HFD. **c**, Normalized ITT at 4 months with HFD. **d–g**, DIO mice were treated with IL-22-Fc or control IgG for 6 weeks (**d–g**) or 4 weeks (**h–j**) ($n = 10$). **d**, Body weight. **e**, Body weight on day 40. **f**, Abdominal photograph of representative mice. **g**, Epididymal fat-pad mass. **h**, Fed and fasted blood glucose. **i**, GTT. **j**, ITT. **e–j**, Age-matched chow-diet-fed mice were used ($n = 10$). Error bars, s.e.m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, unpaired Student's *t*-test (**e, g, h**) or two-way analysis of variance (ANOVA) (**a–d, i, j**). Data shown are representative of three independent experiments.

subsequently developed higher levels of glucose intolerance and insulin resistance within 3 months (Fig. 2a–c and Extended Data Fig. 4d). However, no differences in glucose tolerance and body weight were observed between IL-22-deficient mice and wild-type littermate controls (Extended Data Fig. 4e and data not shown). It is noteworthy that IL-22R1 not only pairs with IL-10R2 to form a functional IL-22 receptor complex, but also couples with the IL-20R2 to form a receptor for IL-20 and IL-24 (ref. 6). The discrepancy in the development of metabolic disorders observed between IL-22- and IL-22R1-deficient mice suggests that other IL-22R1 ligands may play a redundant function in modulating metabolism.

To define the role of IL-22 signalling further in the pathogenesis of metabolic syndrome, we examined the effects of exogenous IL-22-Fc treatment in obese mice. In DIO mice, IL-22-Fc therapy significantly reduced both body weight and epididymal fat-pad mass (Fig. 2d–g), and decreased blood glucose levels under both fed and fasting conditions (Fig. 2h). In addition, 1 month of twice-weekly doses of IL-22-Fc alleviated glucose intolerance and insulin resistance respectively (Fig. 2i, j). Similar results were obtained when mice were concurrently administered with IL-22-Fc from the onset of HFD feeding (Extended Data Fig. 4f–j). In addition, IL-22-Fc treatment significantly reduced parametrial fat-pad mass, lowered blood glucose in fed and fasting conditions, and improved glucose tolerance and insulin sensitivity in *db/db* mice with hyperglycaemia (Extended Data Fig. 5). Continued dosing

of IL-22-Fc was required to maintain the benefit in glucose levels and glucose tolerance both in DIO and in *db/db* mice (data not shown).

In further support of enhanced insulin sensitivity, the serum insulin concentration after IL-22-Fc treatment was significantly reduced in fed and fasted DIO mice (Fig. 3a, b), but not in *db/db* mice (Extended Data Fig. 6a, b). Importantly, improved insulin responsiveness was supported by the increased Akt phosphorylation in muscle, liver, and adipose tissues in IL-22-Fc-treated DIO mice after insulin challenge (Fig. 3c, d). The reduced serum insulin observed after IL-22-Fc treatment was not due to a failure in insulin production. On the contrary, IL-22-Fc treatment restored the ability to produce insulin in response to glucose challenge both in DIO and in *db/db* mice (Fig. 3e and Extended Data Fig. 6c). Consistent with these observations, immunofluorescent staining revealed that IL-22-Fc treatment in *db/db* mice enhanced the islet insulin-staining and the β -cell area per islet but not the numbers and area of islets in pancreas, suggesting an increased insulin storage in the islets (Extended Data Fig. 6d–j). Taken together, our data demonstrate that IL-22-Fc represents a potential therapeutic candidate to normalize blood glucose and alleviate insulin resistance.

IL-22 is canonically associated with antimicrobial immunity and maintenance of gut integrity. Because metabolic syndromes are associated with changes of intestinal microflora, we next determined whether therapeutic effects of IL-22-Fc were due to altering the host microbiome. IL-22-Fc treatment in DIO mice significantly increased the ratio of Bacteroidetes/Firmicutes, which is consistent with previous studies associating decreased phyla ratios with weight gain and obesity (Extended Data Fig. 7a–c)¹⁹. However, IL-22-Fc treatment did not significantly modify microbial compositions at phylum or family levels in wild-type mice according to 16S-rRNA-based metagenomic sequencing (data not shown). This suggests that the change of bacterial composition in obese

mice might not be a direct effect of IL-22, but rather is caused by the amelioration of several aspects of the metabolic syndrome. Indeed, co-housing experiments demonstrated that the therapeutic effects of IL-22 in obese mice could not be transferred to the control mice (Extended Data Fig. 7d, e). These data imply that the metabolic benefits from IL-22-Fc treatment do not simply result from the alteration of the intestinal microbiota, which is in sharp contrast to the microbiota-mediated hyperglycaemia observed in TLR5-deficient mice⁵. However, consistent with its role in preserving epithelial integrity and dampening tissue inflammation, IL-22-Fc treatment resulted in a significant reduction in serum lipopolysaccharide concentration and the expression of adipose-tissue-derived pro-inflammatory genes that are causal with obesity^{20–24} (Fig. 3f–j and Extended Data Fig. 7f, g). Thus, reduced chronic inflammation, but not altered microbiome composition, may contribute to the therapeutic benefits in obesity and metabolic syndrome by IL-22-Fc treatment.

In addition to decreased chronic inflammation, we noticed that IL-22-Fc treatment reduced food intake both in DIO mice (Fig. 4a) and in *db/db* mice (Extended Data Fig. 8a). Enhanced serum levels of PYY, an anorexic gut hormone regulating food intake, was observed in IL-22-Fc treated mice (Fig. 4b and Extended Data Fig. 8b)^{25,26}. Short-term blockade of the PYY pathway using the selective neuropeptide Y (NPY)2 receptor antagonist, BIIIE0246, restored food intake in WT mice treated with IL-22-Fc (Extended Data Fig. 8c)²⁷. However, IL-22-Fc treatment still significantly decreased body weight (DIO mice), blood glucose, and glucose intolerance in pair-feeding studies designed to normalize the food consumption both in DIO and in *db/db* mice (Fig. 4c–e and Extended Data Fig. 8d–f). Conversely, there was no difference in hyperglycaemia and insulin resistance between *ad libitum* and pair-fed control groups despite their different food intakes (Fig. 4c–e and Extended Data Fig. 8d–f). Collectively, these data suggest that although IL-22-Fc reduced food consumption, it does not account solely for all the metabolic benefits seen by IL-22-Fc treatment in obese mice.

Lastly, we examined whether IL-22-Fc therapy had any direct effects on metabolism. IL-22R1 is not only expressed on intestinal epithelial cells but also on several organs involved in metabolism such as liver⁶. Previous studies demonstrated that IL-22 can regulate genes involved in lipogenesis and prevent liver steatosis²⁸. We confirmed that IL-22-Fc treatment directly acted on hepatocytes to induce Stat3 activation (Extended Data Fig. 9a), ameliorated liver steatosis (data not shown), and preserved liver functionality in DIO and *db/db* mice (Extended Data Fig. 9b, c). Hepatic triglyceride and cholesterol levels were also significantly reduced both in DIO and in *db/db* mice treated with IL-22-Fc (Fig. 4f, g and Extended Data Fig. 9d). Additionally, IL-22-Fc directly activated Stat3 in adipose tissue and primary adipocytes (Fig. 4h and data not shown), and promoted the expression of genes involved in triglyceride lipolysis (*Lipe* and *Pnpla2*) and fatty-acid β -oxidation (*Acox1*) (Fig. 4i–k)^{29,30}. Consequently, white adipose tissue (WAT)-associated triglyceride concentration was significantly reduced in obese mice treated with IL-22-Fc (Fig. 4l and Extended Data Fig. 9e). Consistent with these results, serum triglyceride, glycerol, and free-fatty-acid levels were all dramatically decreased in DIO mice after IL-22-Fc treatment (Fig. 4m), supporting a crucial role of IL-22 in modulating lipid metabolism. Interestingly, IL-20 and IL-24 induced similar Stat3 activation and gene expression in adipocytes, but not in hepatocytes (Fig. 4h–k and Extended Data Fig. 9a). IL-22 and IL-20/IL-24 exert distinct functions on other tissues such as intestine owing to the differential expression of the IL-10R2 and IL-20R2 chains (Extended Data Fig. 9f, g)⁹. The partially redundant functions of IL-22 and IL-20/IL-24 pathways in regulating metabolism might explain the different outcomes in IL-22R1- and IL-22-deficient mice fed with HFD. However, *in vivo* administration of either IL-20-Fc or IL-24-Fc did not have similar therapeutic effects compared with IL-22-Fc therapy in obese mice (Extended Data Fig. 9h, i), suggesting that the modulation of lipid metabolism in adipose tissues is not sufficient to provide therapeutic effects in metabolic disease. Taken together, these data support the idea that the multiple downstream

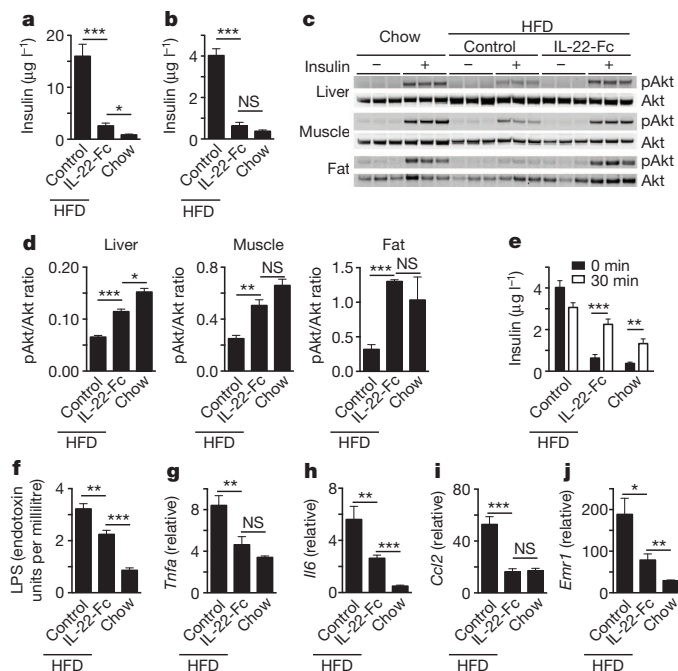


Figure 3 | IL-22 restores insulin sensitivity and dampens chronic inflammation of WAT. DIO mice were treated with IL-22-Fc or control IgG for 8 weeks (a–e), 4 weeks (f), or 6 weeks (g–j). **a**, Fed insulin ($n = 12$). **b**, Fasted insulin ($n = 12$). **c**, Akt phosphorylation in tissues after insulin injection. **d**, Quantification of pAkt relative to total Akt ($n = 3$). **e**, Serum insulin induced by glucose injection ($n = 12$). **f**, Serum lipopolysaccharide ($n = 7$). **g–j**, Gene expression of visceral fat ($n = 10$). **g**, *Tnfa*. **h**, *Il6*. **i**, *Ccl2* (MCP-1). **j**, *Emr1* (F4/80). Age-matched chow-diet-fed mice were used ($n = 10$). Error bars, s.e.m. NS, not significant. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, unpaired Student's *t*-test. Data shown are representative of two independent experiments.

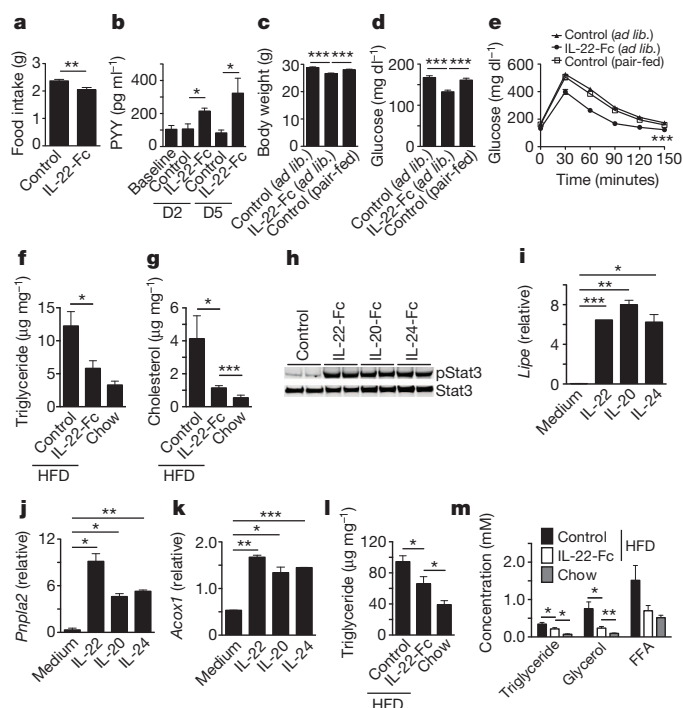


Figure 4 | IL-22 reduces food consumption and regulates lipid metabolism. **a, b,** DIO mice were treated with IL-22-Fc or control IgG. **a,** Overnight food intake on day 2 ($n = 8$). **b,** Serum PYY ($n = 5$). **c–e,** DIO mice ($n = 7$) were fed with food *ad libitum* (*ad lib.*) and treated with IL-22-Fc or control IgG. The food intake of the pair-fed group ($n = 7$) was matched to the food intake of IL-22-Fc treated group, and treated with control IgG. **c,** Body weight on day 27. **d,** Fasted blood glucose on day 28. **e,** GTT on day 28. **f, g, i, m,** DIO mice ($n = 10$) were treated with IL-22-Fc or control IgG for 6 weeks. **f,** Hepatic triglyceride. **g,** Hepatic cholesterol. **h,** *In vivo* Stat3 phosphorylation of WAT from chow-diet-fed mice at 2 h after injection with indicated cytokines. **i–k,** mRNA expression of lipid metabolism genes (**i**, *Lipe*; **j**, *Pnpla2*; **k**, *Acox1*) in primary adipocytes from naive DIO mice stimulated for 8 h with indicated cytokines *in vitro*. **l,** Triglyceride from WAT. **m,** Serum triglyceride, glycerol, and free fatty acids (FFA). **f, g, i, m,** Age-matched chow-diet-fed mice were used ($n = 10$). Error bars, s.e.m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, unpaired Student's *t*-test (**a–d, f, g, i–m**) or two-way ANOVA (**e**). Data shown are representative of two independent experiments.

functions of IL-22 in modulating metabolic pathways are necessary for its therapeutic outcomes.

In summary, although chronic inflammation is well established in promoting obesity and metabolic syndrome, certain leukocyte cytokines, such as IL-22, may have beneficial roles. The IL-22 pathway is indispensable for maintaining epithelial integrity, reducing chronic inflammation, and alleviating metabolic syndromes. Thus, IL-22 can also therapeutically reverse the disease progression through multiple pathways. Further research into its broad biological functions, including mucosal host defence, glucose homeostasis, and lipid metabolism, may help to identify novel therapeutic strategies for the treatment of human metabolic diseases.

METHODS SUMMARY

All animal experiments were approved by the Genentech Institutional Animal Care and Use Committee. For the *C. rodentium* infection study, mice were orally inoculated with 2×10^9 colony-forming units of bacteria. Where indicated, mice were injected with IL-22-Fc or the equivalent amount of control protein three times a week. Colon was harvested on day 4 for analysis of RNA expression. Spleen and liver were harvested on day 10 for bacterial burden analysis. For metabolic analysis, fed blood glucose was measured in the morning, and fasted blood glucose was measured after 16 h overnight fasting. For the glucose tolerance test (GTT) and insulin secretion assays, mice were fasted overnight, and injected intraperitoneally with glucose solution at 1 g kg^{-1} for *db/db* mice and 2 g kg^{-1} for DIO mice. For the insulin tolerance

test (ITT) assay, mice were fasted for 3 h and injected intraperitoneally with insulin solution at one unit per kilogram. For the insulin sensitivity study, mice were fasted overnight, injected intraperitoneally with insulin, and liver, skeletal muscle, and visceral fat were harvested at 10 min after injection for phospho-Akt analysis.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 19 May 2013; accepted 6 June 2014.

Published online 6 August 2014.

1. Tremaroli, V. & Backhed, F. Functional interactions between the gut microbiota and host metabolism. *Nature* **489**, 242–249 (2012).
2. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031 (2006).
3. Gregor, M. F. & Hotamisligil, G. S. Inflammatory mechanisms in obesity. *Annu. Rev. Immunol.* **29**, 415–445 (2011).
4. Kanneganti, T. D. & Dixit, V. D. Immunological complications of obesity. *Nature Immunol.* **13**, 707–712 (2012).
5. Vijay-Kumar, M. *et al.* Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. *Science* **328**, 228–231 (2010).
6. Ouyang, W., Rutz, S., Crellin, N. K., Valdez, P. A. & Hymowitz, S. G. Regulation and functions of the IL-10 family of cytokines in inflammation and disease. *Annu. Rev. Immunol.* **29**, 71–109 (2011).
7. Colonna, M. Interleukin-22-producing natural killer cells and lymphoid tissue inducer-like cells in mucosal immunity. *Immunity* **31**, 15–23 (2009).
8. Myers, M. G., Cowley, M. A. & Munzberg, H. Mechanisms of leptin action and leptin resistance. *Annu. Rev. Physiol.* **70**, 537–556 (2008).
9. Zheng, Y. *et al.* Interleukin-22 mediates early host defense against attaching and effacing bacterial pathogens. *Nature Med.* **14**, 282–289 (2008).
10. Sugimoto, K. *et al.* IL-22 ameliorates intestinal inflammation in a mouse model of ulcerative colitis. *J. Clin. Invest.* **118**, 534–544 (2008).
11. Sonnenberg, G. F. *et al.* Innate lymphoid cells promote anatomical containment of lymphoid-resident commensal bacteria. *Science* **336**, 1321–1325 (2012).
12. Guo, X. *et al.* Leptin signaling in intestinal epithelium mediates resistance to enteric infection by *Entamoeba histolytica*. *Mucosal Immunol.* **4**, 294–303 (2011).
13. Van Maele, L. *et al.* TLR5 signaling stimulates the innate production of IL-17 and IL-22 by CD3^{neg}CD127⁺ immune cells in spleen and mucosa. *J. Immunol.* **185**, 1177–1185 (2010).
14. Procaccini, C., Jirillo, E. & Matarese, G. Leptin as an immunomodulator. *Mol. Aspects Med.* **33**, 35–45 (2012).
15. Lord, G. M. *et al.* Leptin modulates the T-cell immune response and reverses starvation-induced immunosuppression. *Nature* **394**, 897–901 (1998).
16. De Rosa, V. *et al.* A key role of leptin in the control of regulatory T cell proliferation. *Immunity* **26**, 241–255 (2007).
17. Yu, Y. *et al.* Leptin-induced ROR γ t expression in CD4⁺ T cells promotes Th17 responses in systemic lupus erythematosus. *J. Immunol.* **190**, 3054–3058 (2013).
18. Maaser, C. *et al.* Clearance of *Citrobacter rodentium* requires B cells but not secretory immunoglobulin A (IgA) or IgM antibodies. *Infect. Immun.* **72**, 3315–3324 (2004).
19. Tilg, H. & Kaser, A. Gut microbiome, obesity, and metabolic dysfunction. *J. Clin. Invest.* **121**, 2126–2132 (2011).
20. Han, M. S. *et al.* JNK expression by macrophages promotes obesity-induced insulin resistance and inflammation. *Science* **339**, 218–222 (2013).
21. Uysal, K. T., Wiesbrock, S. M., Marino, M. W. & Hotamisligil, G. S. Protection from obesity-induced insulin resistance in mice lacking TNF- α function. *Nature* **389**, 610–614 (1997).
22. Ventre, J. *et al.* Targeted disruption of the tumor necrosis factor- α gene: metabolic consequences in obese and nonobese mice. *Diabetes* **46**, 1526–1531 (1997).
23. Hirosumi, J. *et al.* A central role for JNK in obesity and insulin resistance. *Nature* **420**, 333–336 (2002).
24. Zhang, X. *et al.* Selective inactivation of c-Jun NH₂-terminal kinase in adipose tissue protects against diet-induced obesity and improves insulin sensitivity in both liver and skeletal muscle in mice. *Diabetes* **60**, 486–495 (2011).
25. Batterham, R. L. *et al.* Gut hormone PYY_{3–36} physiologically inhibits food intake. *Nature* **418**, 650–654 (2002).
26. Batterham, R. L. *et al.* Inhibition of food intake in obese subjects by peptide YY_{3–36}. *N. Engl. J. Med.* **349**, 941–948 (2003).
27. Ortiz, A. A. *et al.* A novel long-acting selective neuropeptide Y2 receptor polyethylene glycol-conjugated peptide agonist reduces food intake and body weight and improves glucose metabolism in rodents. *J. Pharmacol. Exp. Ther.* **323**, 692–700 (2007).
28. Yang, L. *et al.* Amelioration of high fat diet induced liver lipogenesis and hepatic steatosis by interleukin-22. *J. Hepatol.* **53**, 339–347 (2010).
29. Odegaard, J. I. *et al.* Macrophage-specific PPAR γ controls alternative activation and improves insulin resistance. *Nature* **447**, 1116–1120 (2007).
30. Zechner, R. *et al.* Fat signals - lipases and lipolysis in lipid metabolism and signaling. *Cell Metab.* **15**, 279–291 (2012).

Acknowledgements We acknowledge S. Haller for her work in pathological analysis of colon samples. We thank F. Chu and J. Eastham-Anderson for performing the immunohistochemical staining and analysing pancreatic cell staining.

Author Contributions W.O. and G.K. devised the project. X.W. and N.O. designed and performed most of the experiments and analyses. P.M. performed western blot and contributed to lipid metabolism studies. G.K., L.K., J.Z.-S., and J.R. contributed to metabolic studies. C.E. and J.L. contributed to flagellin stimulation studies. J.Z. and W.P.L. contributed to ovalbumin/complete Freund's adjuvant stimulation studies.

L.D. did all histology analyses. W.O., X.W., and N.O. prepared the manuscript. N.v.B., P.M., C.E., and G.K. helped to edit the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Readers are welcome to comment on the online version of the paper. The authors declare competing financial interests: details are available in the online version of the paper. Correspondence and requests for materials should be addressed to W.O. (ouyang@gene.com) or G.K. (kolumam@gene.com).

METHODS

Mice. Leptin-receptor-deficient mice (*db/db*; BKS.Cg-*Dock7tm Lepr^{db/+}* +/J or B6.BKS(D)-*Lepr^{db/+}*), leptin-deficient mice (*ob/ob*; B6.Cg-*Lep^{ob/+}*), and their respective lean control mice, as well as HFD mice (C57BL/6J 60%DIO) and the chow-diet-fed control mice, were purchased from The Jackson Laboratory. IL-22-deficient³¹ and IL-22R1-deficient mice (Extended Data Fig. 4a–c) were generated by Lexicon Pharmaceuticals and backcrossed with C57BL/6 strain more than ten times. Where indicated as DIO or HFD, male mice were fed with adjusted calories diet (HFD, containing 60% fat, Harlan) starting at the age of 4–8 weeks old for at least 8 weeks. For metabolism studies, 10- to 16-week-old *db/db* female mice, and 14- to 20-week-old DIO male mice were used. All therapeutic studies with DIO mice were done with age-matched chow-diet-fed male mice. Baseline blood glucose and body weight were measured before treatment. DIO mice were randomized into study groups based on body weight, and *db/db* mice were randomized into study groups based on blood glucose, so that there was no statistically significant difference between each group. During studies, animals presenting severe wounds or that encountered sudden death were excluded from the final analyses. For *in vivo* IL-22 induction and *C. rodentium* infection studies, 5- to 8-week-old female mice of *db/db*, *ob/ob*, and control female mice, or 14- to 18-week-old DIO and chow-diet-fed control male mice, were used. No randomization was done for the infection studies. In each animal study, age- and gender-matched mice were used. The sample size in each study is described in the figure legends. All animal experiments were approved by the Genentech Institutional Animal Care and Use Committee.

***In vivo* stimulation with ovalbumin and flagellin.** To activate CD4⁺ T cells *in vivo*, 100 µg ovalbumin emulsified in complete Freund's adjuvant was injected subcutaneously in the lower back of the animals, and the inguinal lymph nodes were harvested on day 7. To activate TLR5, 3 µg of ultrapure flagellin (InvivoGen) was injected intravenously, and serum samples were harvested at 2 h.

Blood glucose, and serum endotoxin, alanine transaminase, aspartate transaminase, and insulin measurement. Blood glucose was measured by a Contour (Bayer) glucometer for most of the studies and a StatStrip Glucose Xpress (Nova Biomedical) glucometer for Extended Data Fig. 8f. Serum endotoxin was measured by a Limulus Amebocyte Lysate assay kit, QCL-1000 (Lonza), following the manufacturer's instructions. Alanine transaminase and aspartate transaminase were measured by Cholestech LDX (Alere). Serum insulin was measured with an ultra-sensitive mouse insulin ELISA kit (Crystal Chem).

Triglyceride, glycerol, cholesterol, and free-fatty-acid measurement. Lipids were extracted by the method of ref. 32. Serum and intracellular triglyceride was measured using an AdipoSIGHT serum triglyceride assay (Zenbio), and serum glycerol and fatty acids were measured using serum/plasma glycerol and fatty-acid detection kits (Zenbio). Intracellular cholesterol was measured using an Amplex Red cholesterol kit (Life Technologies).

Metabolic analyses. Fed blood glucose was measured at 9:00–10:00 in the morning, and fasted blood glucose was measured after 16 h overnight fasting unless otherwise specified. For the GTT and insulin secretion assays, mice were fasted overnight for about 16 h, and injected intraperitoneally with glucose solution at 1 g kg^{−1} for *db/db* mice and 2 g kg^{−1} for DIO mice. Serum samples for the insulin secretion assay were taken before and after glucose injection. For ITT, mice were fasted for 3 h and injected intraperitoneally with insulin (Humulin NDC 0002-8215-01, Eli Lilly) solution at one unit per kilogram. Blood glucose was measured before and after the injection. For the insulin sensitivity study, mice were fasted overnight and followed with an intraperitoneal injection of insulin (one unit per kilogram). Liver, skeletal muscle, and visceral fat were harvested and snap-frozen in RIPA buffer 10 min after injection for phospho-Akt analysis.

Fc protein therapeutic treatment for metabolic studies. Mice were injected intraperitoneally with IL-22-Fc (IL-22-mIgG2a, 50–100 µg per dose, Genentech, PRO 312045, half-life (*t*_{1/2}) = 3.02 days) or an equivalent amount of isotype control IgG twice a week. Body weight and blood glucose were monitored in the morning two or three times per week. IL-20-Fc (IL-20-mIgG2a, Genentech, PRO355620) and IL-24-Fc (IL-24-mIgG2a, Genentech, PRO370875) were injected at 200 µg per dose, intraperitoneally twice a week. For *in vivo* Stat3 activation, Fc proteins were injected intravenously at 200 µg per dose. Liver and visceral fat tissues were harvested 2 h after injection and immediately snap-frozen in RIPA buffer.

Pair-feeding study. Food consumption was measured daily for the group fed *ad libitum* during the study. The supplied food for the pair-fed group was restricted to match the previous day's food consumption of the group fed *ad libitum*. Correspondingly, the treatment and measurement of the pair-fed group was 1 day after the group fed *ad libitum*. For the restricted food-intake study with DIO mice, 1.8–2.2 g of HFD was given daily. The amount of food was adjusted based on the food intake of the group treated with IL-22-Fc.

PYY measurement and inhibition. Mice were injected intraperitoneally with 50 µg IL-22-Fc on days 0 and 2. On day 4, mice were fasted overnight (16 h) and re-fed for 1 h on day 5. Serum samples were collected on day 2 before treatment and on

day 5 after feeding. All serum samples were mixed with Protease inhibitor (Sigma), DPPIV inhibitor (Millipore), and Pefabloc (Roche) immediately after collection. PYY was measured with a PYY ELISA kit (Abnova) following the manufacturer's instructions. For food intake measurement under PYY inhibition, WT mice were dosed with control IgG or IL-22-Fc (100 µg) on days 0 and 2. Mice were fasted overnight (16 h), then PBS or NYP2 R inhibitor BIIE0246 (40 µg, R&D systems) was injected subcutaneously in the lower back of the animals. Food was given to the animals 45 minutes after injection and then food intake over 4 h of feeding was measured.

Primary adipocyte isolation and stimulation. Primary adipocytes from DIO mice were isolated from visceral fat as described previously³³. Briefly, abdominal fat tissues were harvested, gently homogenized, and digested with collagenase I (Sigma) in adipocyte isolation buffer for 30 min at 37 °C. The mixture was filtrated and washed with PBS by centrifugation. Floating adipocytes were separated and cultured in DMEM (1% FBS) overnight, then stimulated with recombinant proteins (100 ng ml^{−1}; IL-22: 582-ML-010; IL-20: 1024-ML-025; IL-24: 2786-ML-025; R&D systems) for 8 h for gene expression study.

Infection with *C. rodentium* and treatment. *C. rodentium* was cultured overnight and mice were orally inoculated with 2 × 10⁹ colony-forming units of bacteria as described⁹. Where indicated, the mice were injected intramuscularly with IL-22-Fc (150 µg per dose) or control IgG three times per week starting 7 d before infection. For IL-23 induction by hydrodynamic tail vein injection, 10 µg of plasmid coding IL-23 p19 and IL-12/23 p40 or control plasmid were injected as described 3 d before the infection³⁴. IL-23 expression was evaluated by testing serum samples harvested 24 h after the injection with ELISA (eBioscience), and the mice with less than 200 ng ml^{−1} were monitored for survival after infection.

Colon cell isolation. Colons were harvested from the mice infected with *C. rodentium* on day 4 and incubated in HBSS containing 2% FBS, 10 mM HEPES (pH 7.2), 1 mM EDTA and 2 mM DTT with shaking at 230 r.p.m. for 40 min at 37 °C. After vortex for 10 s, the tissue pieces were washed with HBSS containing 2% FBS and 10 mM HEPES (pH 7.2), minced into 1–2 mm, and incubated in HBSS containing 2% FBS, 10 mM HEPES (pH 7.2), 0.1 mg ml^{−1} Liberase TL (Roche), and 0.15 mg ml^{−1} DNase I (Roche) with rotating for 30 min at 37 °C. After terminating the enzyme reaction by adding DMEM containing 10% FBS, the cell suspensions were filtrated, and the cells were washed several times with DMEM. Cells were stimulated with PMA (10 ng ml^{−1}) and ionomycin (500 ng ml^{−1}) for 4 h in the presence of Brefeldin A, and IL-22 expression was analysed by intracellular staining as described below.

Histological evaluation. Immunohistochemistry for insulin and glucagon was performed on formalin-fixed paraffin-embedded pancreas tissues as previously reported³⁵ using rabbit anti-glucagon (Cell Signaling Technologies, 2760) with Alexa Fluor 555-conjugated goat anti-rabbit secondary antibody, or guinea-pig anti-insulin (DAKO, A0564) with Alexa Fluor 647-conjugated goat anti-guinea-pig secondary antibody. The percentage insulin area per islet area was calculated by dividing the insulin positive area by the islet area minus the nuclear area. Histological analysis of colons from mice infected with *C. rodentium* was performed as reported previously³⁴, and scored for epithelial changes (proliferation, blebbing, enterocyte shedding), inflammation, and mucosal thickening. Clinical scores were determined for four anatomical regions (proximal, middle colon, distal colon, and rectum) on a scale from 0 to 5 (0 = normal, 5 = severe disease). Regional scores were summed to get a final disease severity score for the colon for each animal. Histological analyses were performed blindly.

RNA isolation and real-time PCR. Colons and visceral fat tissues were harvested and processed with gentle MACS (Miltenyi Biotec), and mRNA was isolated with an RNeasy mini plus kit (Qiagen) or Trizol and PureLink RNA kit (Life Technologies). RNA was directly analysed by One Step RT-PCR (Qiagen), or complementary DNA (cDNA) was synthesized from RNA samples with a SuperScript VILO cDNA synthesis kit (Life Technologies) followed by real-time PCR with iTaq Universal Probes Supermix (Bio-Rad). ABI7500 (Applied Biosystems) was used for real-time PCR analysis as previously reported³⁴. Results were normalized to those of the control housekeeping gene *Rpl19* (encoding ribosomal protein L19) and are reported as 1,000 × 2^{−ΔCT}. The primer and probe sequences for *Il22* and *Reg3b* were reported previously³⁴. Primer/probe sets for *Tnfr1*, *Il6*, *Il1β*, *Ccl2*, *Emr1*, *Il23a*, *Il12b* and *Il10rb* were purchased from Life Technologies. Other gene expression analyses were done with the following primer/probe sets: *Il22ra1*, 5'-AGG TCC ATT CAG ATG CTG GT-3', 5'-TAG GTG TGG TTG ACG TGG AG-3' and 5'-FAM-CCA CCC CAC ACT CAC ACC GG-TAMRA-3'. *Il20rb*, 5'-GGT TTT CTC ACG GAT GCA GTG-3', 5'-TGC TTC ATG TTG GTT GAC TGT ACA-3', and 5'-FAM-CCG TTT TAC CTG CCC CTC AGA ACC TC-TAMRA-3'. *Acox1*, 5'-CGC ACA TCT TGG ATG GTA GT-3', 5'-GGC TTC GAG TGA GGA AGT TAT AG-3', and 5'-FAM-CGC CGT CGA GAA ATC GAG AAC TTG A-TAMRA-3'. *Lipe*, 5'-CAT CAA CCA CTG TGA GGG TAA G-3', 5'-AAG GGA GGT GAG ATG GTA ACT-3', and 5'-FAM-AGA TGC CAC TCA CCT CTG ATC CCA-TAMRA-3'. *Pnpla2*, 5'-CTT TGT GGC ACA GAC CTC TAA-3', 5'-GGG AGT AGT CGA

TGG AGA AGA TA-3', and 5'-FAM-ATC CCA CGA GCC CTA CCT CAA GAA-TAMRA-3'.

Faecal DNA extraction and analysis. Genomic DNA from faecal samples was purified using a PowerSoil DNA isolation kit (Mo Bio Laboratories). For each sample, 16S rRNA genes were amplified with a set of primers specific to Bacteroidetes and Firmicutes with SYBR Green real-time PCR Master Mix (Life Technologies). The primers for Bacteroidetes were 5'-GGT TCT GAG AGG AGG TCC C-3' and 5'-GCT GCC TCC CGT AGG AGT-3'³⁶. The primers for Firmicutes were 5'-GGA GYA TGT GGT TTA ATT CGA AGC A-3' and 5'-AGC TGA CGA CAA CCA TGC AC-3'³⁷.

Flow cytometry and ELISA. Lymphocytes purified from draining lymph nodes and colons were stained for cytokines as previously described³¹ using phycoerythrin (PE)-conjugated anti-IL-22 (1H8PWSR, eBioscience, 12-7221-82), and fluorescein isothiocyanate (FITC)-conjugated anti-IL-17A (17B7, eBioscience, 11-7177-81), allophycocyanin (APC)-conjugated IFN- γ (XMG1.2, eBioscience, 17-7311-81), anti-Foxp3 (FJK-16 s, eBioscience, 17-5773-82). Colon cells were stained with a LIVE/DEAD fixable Aqua Dead cell stain kit (Life Technologies) according to the manufacturer's protocol, and stained surface markers for eFluor450-conjugated anti-CD45 (30-F11, eBioscience, 48-0451-82), allophycocyanin-indotricarbocyanine (APC-Cy7)-conjugated anti-CD3 ϵ (145-2C11, BD, 557596), FITC-conjugated anti-TCR $\gamma\delta$ (GL3, BD, 553177), anti-CD8 α (53-6.7, BD, 553071), anti-CD11c (HL3, BD, 553801), anti-NK1.1 (PK136, eBioscience, 11-5941-85), and anti-B220 (RA3-6B2, eBioscience, 11-0452-85), phycoerythrin-indotricarbocyanine (PE-Cy5)-conjugated anti-IL-7R α (A7R34, eBioscience, 15-1271-81), Alexa Fluor 700-conjugated anti-CD4 (GK1.5, eBioscience, 56-0041-82), phycoerythrin-indotricarbocyanine (PE-Cy7)-conjugated anti-c-Kit (2B8, eBioscience, 25-1171-82), then fixed with 4% PFA. The cells were made permeable with Permeabilization Buffer (eBioscience), then stained with anti-IL-22. Stained cells were analysed on an LSR II (BD Biosciences) with FlowJo software (TreeStar). ILCs were defined as CD45⁺, lineage-negative (CD3⁻TCR $\gamma\delta$ ⁻CD8 α ⁻CD11c⁻B220⁻NK1.1⁻), IL7R α ⁺, and c-Kit⁺. ELISA for IL-22 was performed as previously described³¹ using monoclonal anti-IL-22 antibody (20E5 and 14B7, Genentech).

Western blot. Liver and muscle samples from chow-diet-fed mice were lysed in standard RIPA buffer (150 mM NaCl, 50 mM Tris, 1% IGEPAL, 0.5% sodium-deoxycholate, 0.1% SDS). WAT samples and livers from obese mice were processed in RIPA buffer supplemented with 1% Triton-X 100. Proteins were quantified with BCA assay (Pierce). Equal amounts of protein were electrophoresed on Bolt 4–12% Bis-Tris plus gels (Life Technologies) and transferred onto nitrocellulose

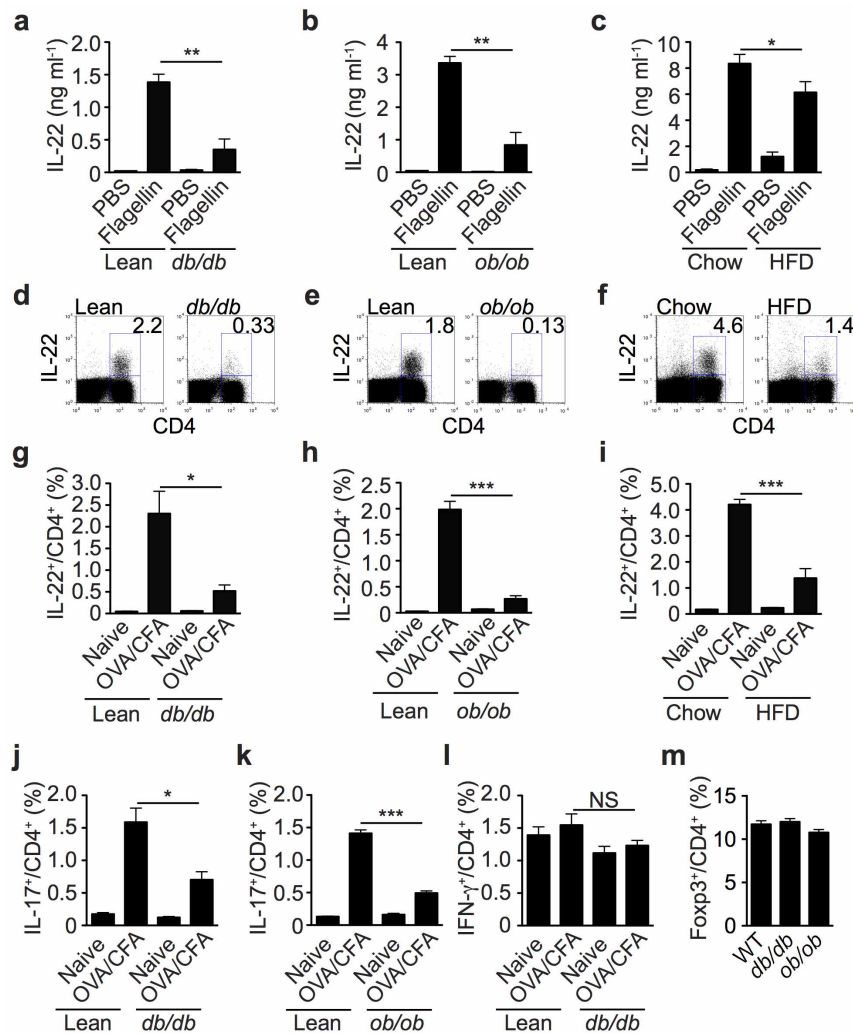
membranes. Western blots were analysed using an Odyssey Imager (Licor) according to the manufacturer's instructions. The quantification was done with Image Studio Lite (Licor). The following antibodies (Cell Signaling Technology) were used: pStat3(Y705)(3E2) 9138s, Stat3(79D7) 4904p, pAkt(S473)(193H12) 4058s, Akt 9272s.

Bacterial burden analysis. The spleens, livers, and colons were harvested, weighted, and homogenized in 0.1% NP40/PBS in a C-tube with gentleMACS (Miltenyi Biotec). Serially diluted homogenates were plated on MacConkey agar (Remel), and *C. rodentium* colonies were identified as pink colonies after overnight incubation at 37 °C.

Detection of anti-*C. rodentium* antibody in serum. The serum samples were harvested on day 10 after the infection. ELISA plate was coated with heat-killed *C. rodentium* or with a goat anti-mouse Ig-capturing antibody. Coated plate was washed with washing buffer (0.05% Tween 20 in PBS), blocked for 2 h with blocking buffer (0.5% BSA, 15 p.p.m. Proclin in PBS), and washed before the addition of serially diluted standard mouse monoclonal IgG (SouthernBiotech), or serum samples. After 2 h incubation at room temperature, the plate was washed and the Ig was detected with goat anti-mouse IgG conjugated with horseradish peroxidase (SouthernBiotech), diluted 1/4,000 in assay diluent (0.5% BSA, 0.05% Tween 20, 15 p.p.m. Proclin in PBS), and incubated for 2 h at room temperature. After washing, TMB peroxidase substrate (Sigma-Aldrich) was added to each well. Absorbance was read at 650 nm on a microplate reader (Molecular Devices).

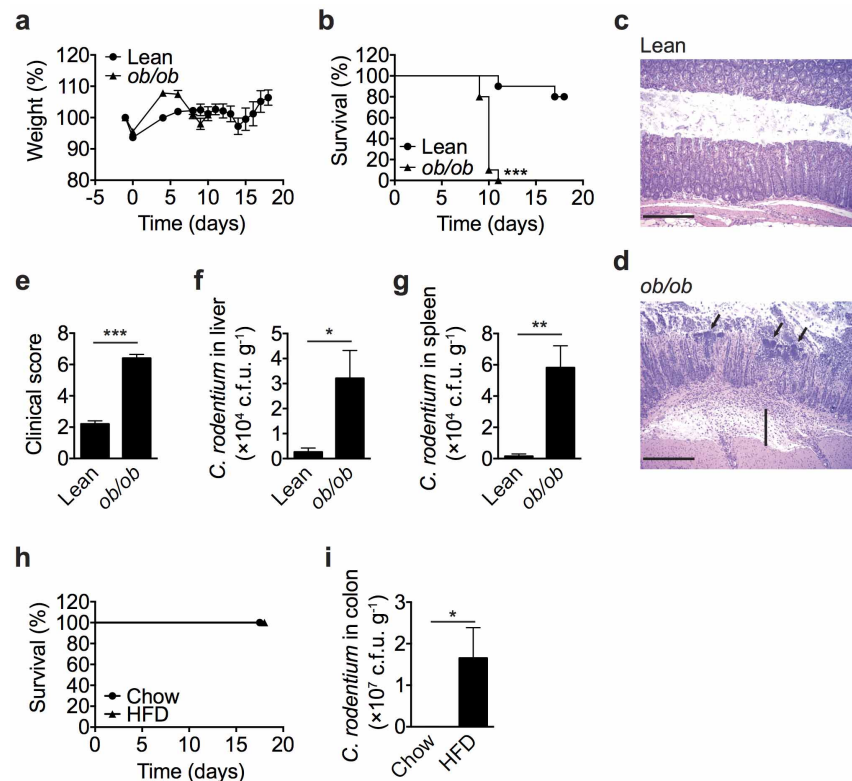
Statistical analysis. All statistical analysis used a two-tailed unpaired Student's *t*-test, two-way ANOVA or log rank test. *P* < 0.05 was considered statistically significant.

31. Zheng, Y. *et al.* Interleukin-22, a Th17 cytokine, mediates IL-23-induced dermal inflammation and acanthosis. *Nature* **445**, 648–651 (2007).
32. Bligh, E. G. & Dyer, W. J. A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* **37**, 911–917 (1959).
33. Rosen, E. D. *et al.* PPAR gamma is required for the differentiation of adipose tissue in vivo and in vitro. *Mol. Cell* **4**, 611–617 (1999).
34. Ota, N. *et al.* IL-22 bridges the lymphotoxin pathway with the maintenance of colonic lymphoid structures during infection with *Citrobacter rodentium*. *Nature Immunol.* **12**, 941–948 (2011).
35. Wu, A. L. *et al.* Amelioration of type 2 diabetes by antibody-mediated activation of fibroblast growth factor receptor 1. *Sci. Translat. Med.* **3**, 113ra126 (2011).
36. Barman, M. *et al.* Enteric salmonellosis disrupts the microbial ecology of the murine gastrointestinal tract. *Infect. Immun.* **76**, 907–915 (2008).
37. Sham, H. P. *et al.* SIGIRR, a negative regulator of TLR/IL-1R signalling promotes microbiota dependent resistance to colonization by enteric bacterial pathogens. *PLoS Pathog.* **9**, e1003539 (2013).



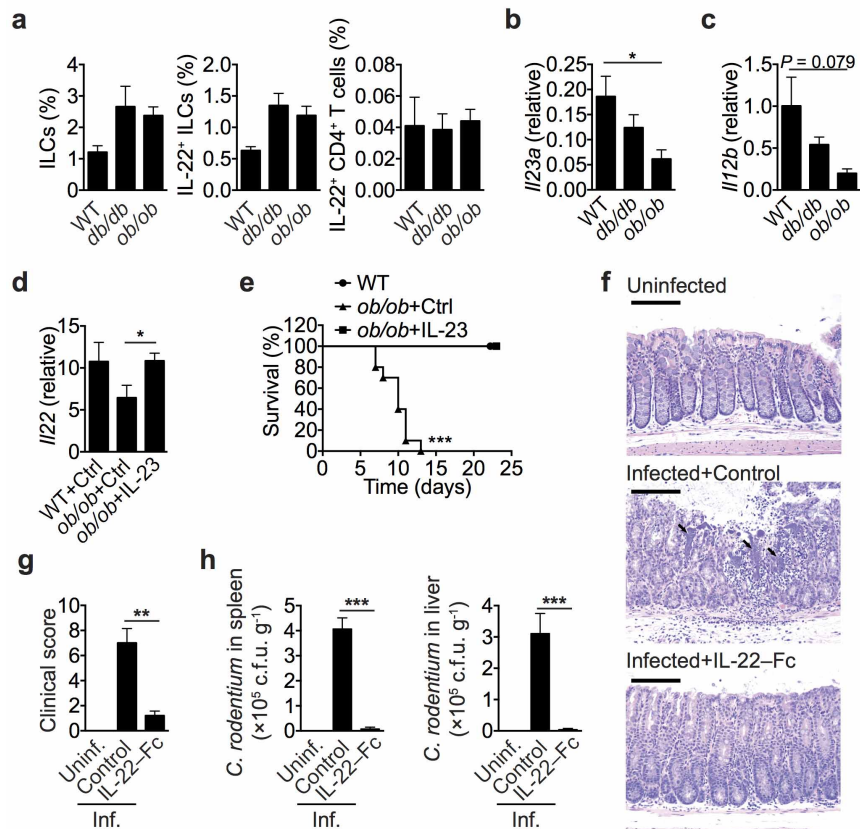
Extended Data Figure 1 | Obese mice have defective IL-22 and IL-17A production. **a–c**, ELISA of IL-22 in serum 2 h after flagellin injection from *db/db* and lean control mice (**a**), *ob/ob* and lean control mice (**b**), and HFD and chow-diet-fed mice (**c**) ($n = 4$). **d–l**, Mice were immunized with ovalbumin/complete Freund's adjuvant. Lymphocytes in draining lymph nodes were analysed for cytokine expression on day 7 ($n = 4$). Flow cytometry analysis of IL-22⁺ lymphocytes (**d–f**), percentage of IL-22⁺ cells (**g–i**), IL-17A⁺ cells (**j, k**), and IFN-γ⁺ cells (**l**) within the CD4⁺ population in *db/db* and lean

controls (**d, g, j, l**), *ob/ob* and lean controls (**e, h, k**), and HFD and chow-diet-fed mice (**f, i**). The number at the top right-hand corner of each FACS plot is the percentage of IL-22⁺ cells within CD4⁺ cells (**d–f**). **m**, Percentage of Foxp3⁺ cells within CD4⁺ T cells from spleens of naive WT, *db/db*, and *ob/ob* mice ($n = 5$). Error bars, s.e.m. NS, not significant. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, unpaired Student's t -test. Data shown are representative of three (**d, e, g, h, j–l**) or two (**a–c, f, i, m**) independent experiments.



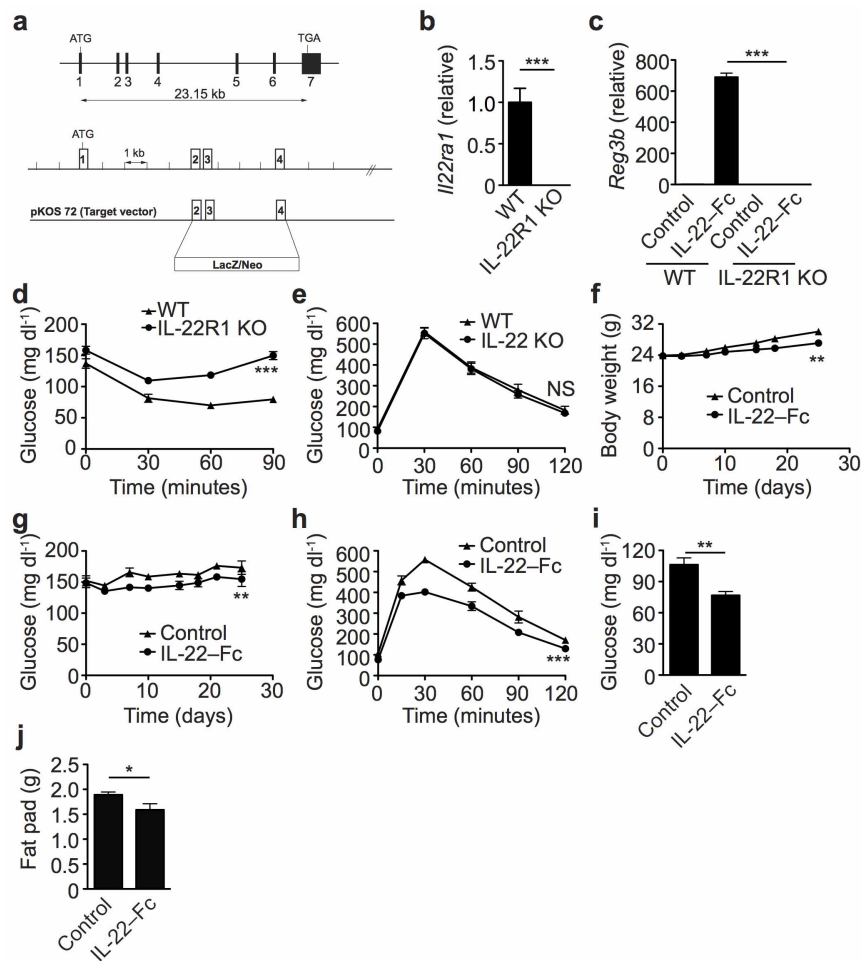
Extended Data Figure 2 | Susceptibility of *ob/ob* and DIO mice to *C. rodentium* infection. **a, b**, Body weight (**a**) and survival (**b**) of *ob/ob* and lean control mice ($n = 10$) infected with *C. rodentium*. **c, d**, Colon histology of lean control (**c**) or *ob/ob* (**d**) mice on day 8 after infection. Arrows, bacterial colonies; vertical bar, submucosal oedema; horizontal bar, 200 μ m. **e**, Clinical score determined by colon histology ($n = 5$). **f, g**, Bacterial burden in liver

(**f**) and spleen (**g**) on day 7 ($n = 5$). **h**, Survival of mice fed with HFD or chow diet infected with *C. rodentium* ($n = 10$). **i**, Bacterial burden in colon on day 18 ($n = 5$). Error bars, s.e.m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, Log rank test (**b**) or unpaired Student's *t*-test (**e–g, i**). Data shown are representative of two independent experiments.



Extended Data Figure 3 | Reduced IL-23 induction contributes to the compromised IL-22 expression and host defence during *C. rodentium* infection in obese mice. **a–c**, WT, *db/db*, and *ob/ob* mice on day 4 after infection. **a**, Percentage of ILCs, IL-22⁺ ILCs, and IL-22⁺ CD4⁺ T cells within leukocytes in colons after *ex vivo* stimulation ($n = 3$). **b**, **c**, Relative *Il23a* (**b**) and *Il12b* (**c**) mRNA expression in colons ($n = 5$). **d**, **e**, WT and *ob/ob* mice were infected with *C. rodentium* and the hydrodynamic tail vein injected with either a control or IL-23-encoding plasmid. **d**, Relative *Il22* mRNA expression on day 4 ($n = 10$ for WT, $n = 6$ for *ob/ob* + Ctrl and *ob/ob* + IL-23). **e**, Survival

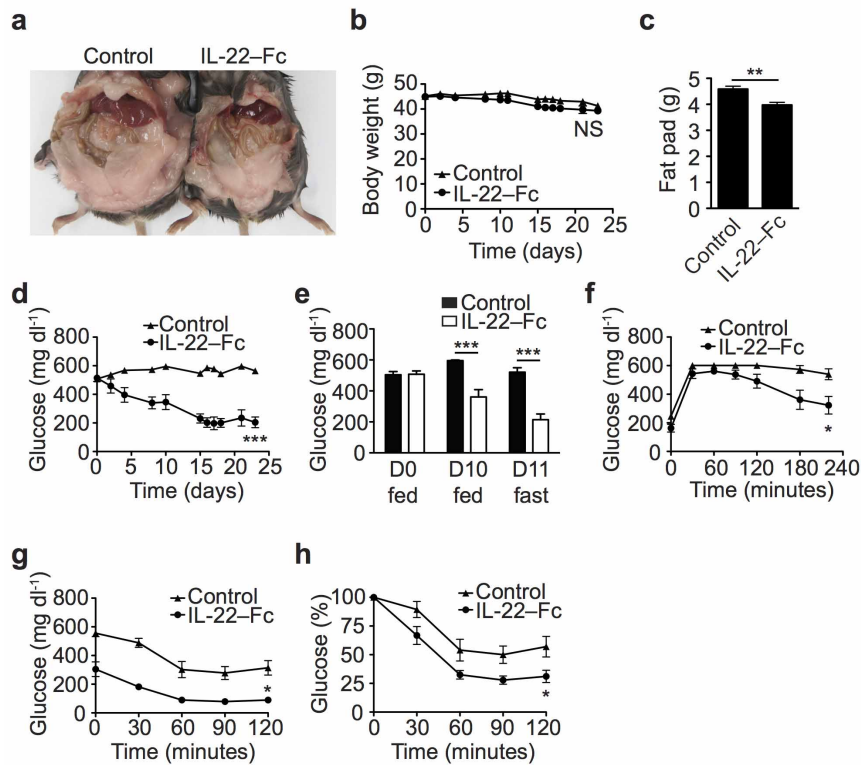
($n = 10$ for WT and *ob/ob* + Ctrl, $n = 8$ for *ob/ob* + IL-23). **f–h**, *db/db* mice were infected with *C. rodentium* and treated with control IgG or IL-22-Fc ($n = 5$ for uninfected and IL-22-Fc, $n = 4$ for control). **f**, Colon histology on day 6. Arrows, bacterial colonies; horizontal bar, 100 μ m. **g**, Clinical score determined by colon histology. **h**, Bacterial burden in spleen (left) and liver (right) on day 8. Error bars, s.e.m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, unpaired Student's *t*-test (**b–d**, **g**, **h**) or log rank test (**e**). Data shown are representative of two independent experiments.



Extended Data Figure 4 | Role of IL-22 pathway in metabolic disorders.

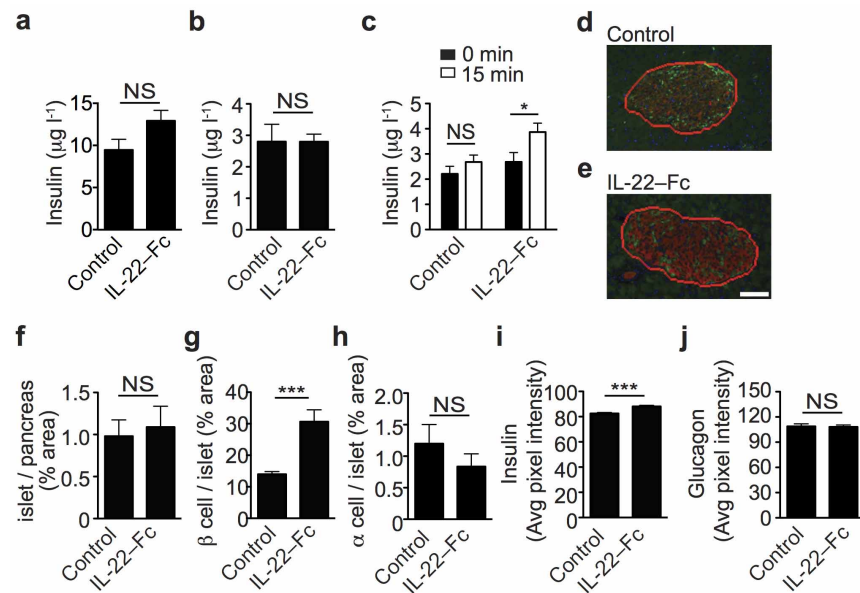
a, Strategy for generation of IL-22R1 KO mice. **b**, *Il22ra1* mRNA expression in colons from IL-22R1 KO and WT littermates ($n = 5$). **c**, *Reg3b* mRNA expression in colons from IL-22R1 KO and WT littermates 2 days after an injection with IL-22-Fc or control IgG ($n = 5$). **d**, **e**, IL-22R1 KO, IL-22 KO ($n = 13$), and WT littermate ($n = 8$) male mice were fed with HFD as in Fig. 2a–c. **d**, ITT of IL-22R1 KO mice as in Fig. 2c. **e**, GTT of IL-22 KO mice at

3 months with HFD. **f–j**, Eight-week-old C57BL/6 mice were concurrently fed with HFD and treated with IL-22-Fc or control IgG ($n = 7$). **f**, Body weight. **g**, Blood glucose. **h**, GTT on day 23. **i**, Fasted blood glucose on day 23. **j**, Epididymal fat-pad mass on day 25. Error bars, s.e.m.; NS, not significant. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, unpaired Student's *t*-test (**b**, **c**, **i**, **j**) or two-way ANOVA (**d–h**). Data shown are representative of three independent experiments.



Extended Data Figure 5 | IL-22 ameliorates metabolic disorders in diabetic *db/db* mice. *db/db* mice were treated with IL-22-Fc or control IgG for 3 weeks ($n = 7$). **a**, Abdominal photograph of representative mice. **b**, Body weight. **c**, Parametrial fat-pad mass. **d**, Blood glucose. **e**, Fed and fasted blood glucose.

f, GTT. **g**, ITT. **h**, Normalized ITT. Error bars, s.e.m.; NS, not significant. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, two-way ANOVA (**b**, **d**, **f**–**h**) or unpaired Student's *t*-test (**c**, **e**). Data shown are representative of five independent experiments.



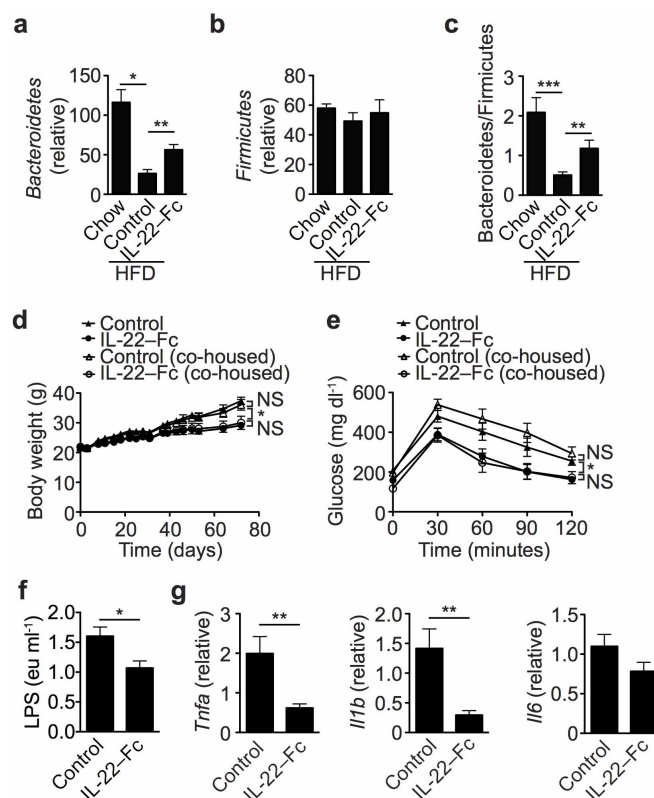
Extended Data Figure 6 | IL-22 restores insulin homeostasis in *db/db* mice.

a–c, *db/db* mice were treated with IL-22-Fc or control IgG for 3 weeks.

a, b, Serum insulin of fed (**a**) and 16 h fasted (**b**) mice. **c**, Insulin secretion in response to glucose injection after fasting for 16 h ($n = 8$). **d–j**, Pancreases were harvested and stained for insulin and glucagon ($n = 7$).

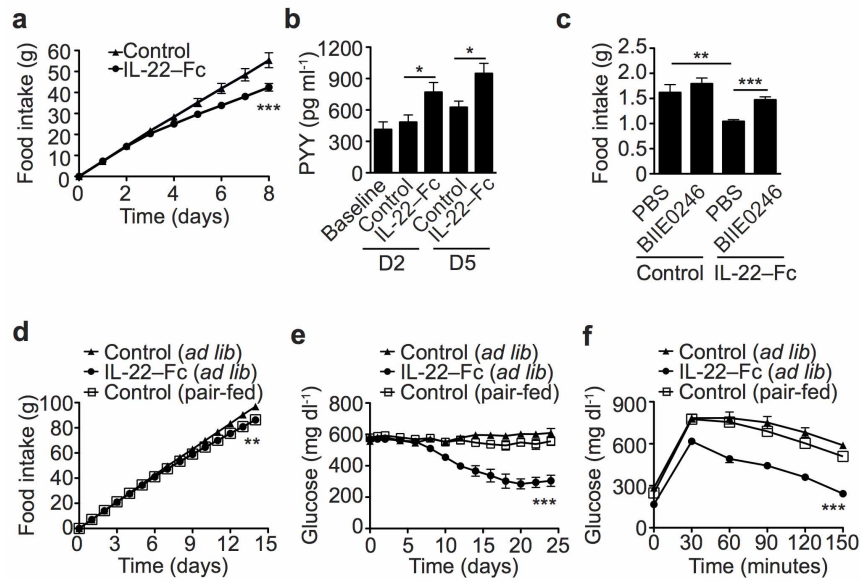
d, e, Immunohistochemistry of pancreas islet from *db/db* mice treated with control IgG (**d**) or IL-22-Fc (**e**). Green, glucagon; red, insulin. The area

surrounded by the red line shows islet area. Scale bar, 50 μm . **f**, Percentage of islet area within the total pancreas area. **g**, Percentage of β -cell area within the total islet area. **h**, Percentage of α -cell area within the total islet area. **i**, Average insulin staining intensity. **j**, Average glucagon staining intensity. Error bars, s.e.m.; NS, not significant. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, unpaired Student's *t*-test. Data shown are representative of two independent experiments.



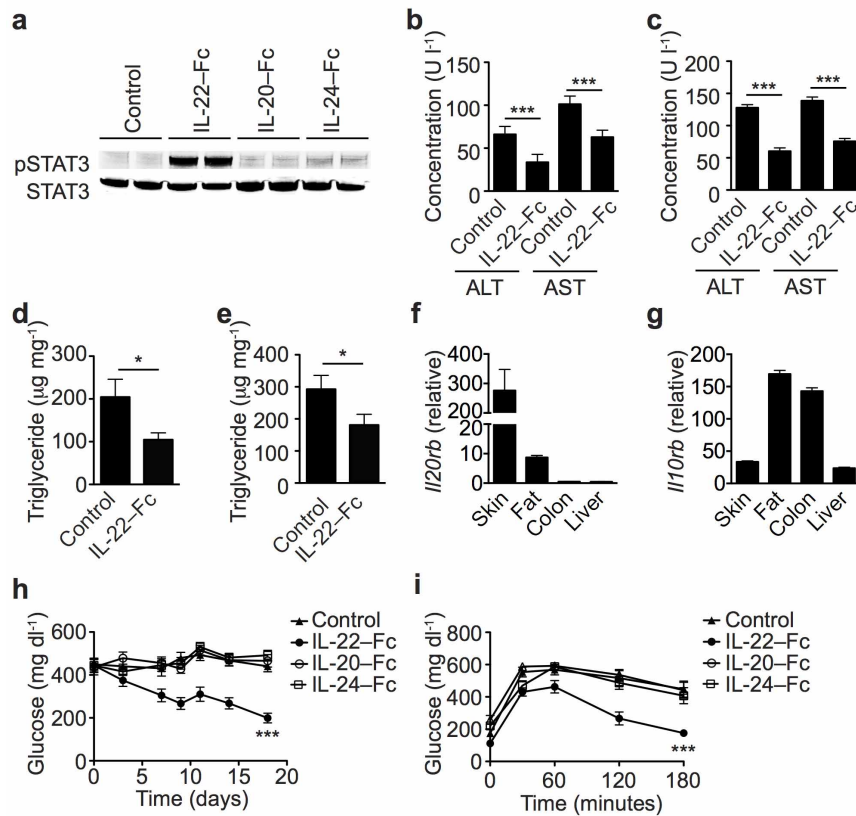
Extended Data Figure 7 | Role of IL-22-Fc in modulation of intestinal microbiota and chronic inflammation. **a–c**, DIO mice were treated with IL-22-Fc or control IgG for 6 weeks ($n = 8$). Faecal DNA was analysed by real-time PCR. **a**, Relative levels of Bacteroidetes 16S rRNA. **b**, Relative levels of Firmicutes 16S rRNA. **c**, Ratio of Bacteroidetes to Firmicutes. **d**, **e**, DIO mice were treated with IL-22-Fc or control IgG for 10 weeks ($n = 5$). One group of IL-22-Fc-treated mice were co-housed with control treated mice ($n = 5$).

d, Body weight. **e**, GTT. **f**, **g**, *db/db* mice were treated with IL-22-Fc or control IgG for 2 weeks ($n = 10$). **f**, Serum lipopolysaccharide. **g**, Real-time PCR analyses of *Tnfa*, *Il1b*, and *Il6* mRNA expression in WAT. Error bars, s.e.m.; NS, not significant. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, unpaired Student's *t*-test (**a**, **c**, **f**, **g**) or two-way ANOVA (**d**, **e**). Data shown are representative of two independent experiments.



Extended Data Figure 8 | The regulation of glucose homeostasis by IL-22-Fc is independent of its function in limiting food intake. **a**, Accumulative food intake in the first week of the *db/db* mice in Extended Data Fig. 5. **b**, Serum PYY in *db/db* mice treated with IL-22-Fc or control IgG ($n = 5$). **c**, Four-hour food intake of IL-22-Fc or control IgG-dosed C57BL/6 male mice in a fast and re-fed study with a single injection of NPY2 R inhibitor, BIIE0246, or PBS ($n = 5$). **d–f**, Two groups of *db/db* mice ($n = 6$) were fed with food *ad libitum*

and treated with IL-22-Fc or control IgG for 4 weeks. One group of *db/db* mice ($n = 6$) was fed with restricted food that matched the food intake of IL-22-Fc treated group, and treated with control IgG. **d**, Accumulative food intake of the first 2 weeks. **e**, Blood glucose. **f**, GTT. Error bars, s.e.m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, two-way ANOVA (**a**, **d–f**) or unpaired Student's *t*-test (**b**, **c**). Data shown are representative of two independent experiments.



Extended Data Figure 9 | IL-22-Fc, but not IL-20-Fc or IL-24-Fc, improves metabolic syndrome. **a**, Stat3 phosphorylation in liver from WT mice treated with Fc proteins (200 µg) for 2 h *in vivo*. **b**, **c**, Liver enzymes of DIO mice (**b**) and *db/db* mice (**c**) treated with IL-22-Fc or control IgG for 4 weeks ($n = 5$). **d**, **e**, *db/db* mice were treated with IL-22-Fc ($n = 9$) or control IgG ($n = 6$) for

4 weeks. **d**, Hepatic triglyceride. **e**, WAT triglyceride. **f**, **g**, *Il20rb* (**f**) and *Il10rb* (**g**) mRNA expression in indicated tissues of WT mice. **h**, **i**, *db/db* mice were treated with Fc proteins for 2 weeks ($n = 10$). **h**, Blood glucose. **i**, GTT. Error bars, s.e.m. * $P < 0.05$, *** $P < 0.001$, unpaired Student's *t*-test (**b–e**) or two-way ANOVA (**h**, **i**).

HSP70 sequestration by free α -globin promotes ineffective erythropoiesis in β -thalassaemia

Jean-Benoît Arlet^{1,2,3,4*}, Jean-Antoine Ribeil^{1,3,4,5*}, Flavia Guillem^{1,3,4}, Olivier Negre⁶, Adonis Hazoume^{7,8}, Guillaume Marcion^{7,8}, Yves Beuzard⁶, Michaël Dussiot^{1,3,4,9}, Ivan Cruz Moura^{1,3,4,9,10}, Samuel Demarest¹¹, Isaure Chauvot de Beauchêne^{11,12}, Zakia Belaid-Choucair^{1,3,4}, Margaux Sevin^{7,8}, Thiago Trovati Maciel^{1,3,4,9,10}, Christian Auclair^{11,12}, Philippe Leboulch^{6,13}, Stany Chretien⁶, Luba Tchertanov^{11,12}, Véronique Baudin-Creuzat¹⁴, Renaud Seigneure⁸, Michaela Fontenay^{4,15}, Carmen Garrido^{7,8,16§}, Olivier Hermine^{1,3,4,17§} & Geneviève Courtois^{1,3,4§}

β -Thalassaemia major (β -TM) is an inherited haemoglobinopathy caused by a quantitative defect in the synthesis of β -globin chains of haemoglobin, leading to the accumulation of free α -globin chains that form toxic aggregates^{1,2}. Despite extensive knowledge of the molecular defects causing β -TM, little is known of the mechanisms responsible for the ineffective erythropoiesis observed in the condition, which is characterized by accelerated erythroid differentiation, maturation arrest and apoptosis at the polychromatophilic stage^{3–6}. We have previously demonstrated that normal human erythroid maturation requires a transient activation of caspase-3 at the later stages of maturation⁷. Although erythroid transcription factor GATA-1, the master transcriptional factor of erythropoiesis, is a caspase-3 target, it is not cleaved during erythroid differentiation. We have shown that, in human erythroblasts, the chaperone heat shock protein70 (HSP70) is constitutively expressed and, at later stages of maturation, translocates into the nucleus and protects GATA-1 from caspase-3 cleavage⁸. The primary role of this ubiquitous chaperone is to participate in the refolding of proteins denatured by cytoplasmic stress, thus preventing their aggregation⁹. Here we show *in vitro* that during the maturation of human β -TM erythroblasts, HSP70 interacts directly with free α -globin chains. As a consequence, HSP70 is sequestered in the cytoplasm and GATA-1 is no longer protected, resulting in end-stage maturation arrest and apoptosis. Transduction of a nuclear-targeted HSP70 mutant or a caspase-3-uncleavable GATA-1 mutant restores terminal maturation of β -TM erythroblasts, which may provide a rationale for new targeted therapies of β -TM.

To investigate the hypothesis that HSP70 can be sequestered in the cytoplasm of mature β -TM erythroblasts by binding to free α -globin chains of haemoglobin or aggregates, we analysed its subcellular localization in bone marrow samples from adult β -TM patients and healthy donors by confocal microscopy ($n = 3$). In contrast to normal erythroblasts, in β -TM mature haemoglobinized erythroblasts, HSP70 was mainly localized in the cytoplasm and GATA-1 was poorly expressed in the nucleus (Fig. 1).

To explore the role of HSP70 in β -TM ineffective erythropoiesis, we performed an *in vitro* two-phase liquid culture, allowing proliferation,

survival and erythroid differentiation of β -TM or control CD34⁺ progenitors towards the formation of acidophilic erythroblasts and reticulocytes. During the second phase of amplification (day 8), corresponding to erythroid terminal differentiation, differentiation in the β -TM group was accelerated. This was characterized by a higher percentage of polychromatophilic cells ($26.2 \pm 8.4\%$ versus $9.0 \pm 2.7\%$; $P = 0.003$) (Fig. 2a), accelerated downregulation of stem cell growth factor receptor (CD117, also known as c-Kit), an early erythroid marker (mean fluorescence intensity (MFI) 25.7 ± 17 versus 115.1 ± 47.1 ; $P = 0.0001$), and upregulation of glycophorin-A (MFI 243.8 ± 87.8 versus 178.9 ± 56 ; $P = 0.04$) and band 3 (MFI 118.4 ± 37.9 versus 60.6 ± 15.2 ; $P = 0.03$) (Extended Data Fig. 1). At the time of intense haemoglobinization (days 8–10), apoptosis was increased (at day 10, $38.2\% \pm 15.1$ versus $18.5\% \pm 8.7$; $P = 0.01$) (Fig. 2b) and terminal maturation was arrested at the polychromatophilic stage in β -TM cells. To quantify this maturation arrest, we defined an index of terminal maturation (TMI) as the number of (acidophilic cells + reticulocytes per slide) $\times 100$ divided by the number of polychromatophilic cells per slide. At day 8, TMI was significantly decreased in β -TM cells: 16% (interquartile range (IQR) 8.8–27.8%) compared to 37.6% in controls (IQR 24.4–70.7%) ($P = 0.009$) (Fig. 2a). Taken together, this cell-culture system reproduced the characteristics of β -TM ineffective erythropoiesis, namely accelerated differentiation, maturation arrest and death of mature haemoglobinized erythroblasts^{3–6}. Next, we analysed the subcellular localization of HSP70 at several time intervals in differentiating erythroblasts from β -TM patients and healthy donors by confocal microscopy. In agreement with our observations in bone marrow, HSP70 was detected in the nucleus of control erythroblasts but was absent or only weakly expressed in mature haemoglobinized β -TM cells. Thus, at day 8, the ratio of cytoplasmic to nuclear HSP70 MFI in β -TM erythroblasts was significantly increased, with a median ratio of 2.3 (IQR 1.6–3) compared to 1.1 in controls (IQR 0.7–1.6) ($P < 0.0001$) (Fig. 2c). As a result, GATA-1 was poorly expressed in the nucleus of mature haemoglobinized β -TM erythroblasts (Fig. 2c), thus supporting our hypothesis. To further understand the link between haemoglobinization, HSP70 localization and decrease in GATA-1 expression, we studied changes in the expression

¹Laboratoire INSERM, unité mixte de recherche 1163, centre national de la recherche scientifique (CNRS) équipe de recherche labellisée 8254, 24 Boulevard de Montparnasse, 75015 Paris, France.

²Service de Médecine Interne, Faculté de médecine Paris Descartes, Sorbonne Paris-Cité et Assistance publique – Hôpitaux de Paris, Hôpital Européen Georges Pompidou, 15 rue Leblanc 75908 Paris, France.

³Paris Descartes-Sorbonne Paris Cité University, Imagine Institute, Assistance publique – Hôpitaux de Paris, Hôpital Necker, 24 Boulevard de Montparnasse, 75015 Paris, France. ⁴Laboratory of Excellence GR-Ex, 75015 Paris, France. ⁵Département de Biothérapie, Faculté de médecine Paris Descartes, Sorbonne Paris-Cité et Assistance publique – Hôpitaux de Paris, Hôpital Necker, 149 rue de Sèvres 75015 Paris, France. ⁶Commissariat à l'énergie atomique (CEA), Institute of Emerging Diseases and Innovative Therapies (IMETI), 18 Route du Panorama, 92260 Fontenay-aux-Roses, France.

⁷INSERM, unité mixte de recherche 866, Equipe labellisée Ligue contre le Cancer and Association pour la Recherche contre le Cancer, and Laboratoire d'Excellence Lipoprotéines et santé (LipSTIC), 21033 Dijon, France. ⁸University of Burgundy, Faculty of Medicine and Pharmacy, 7 boulevard Jeanne d'Arc, 21033 Dijon, France. ⁹INSERM, unité mixte de recherche 699, Hôpital Bichat, 46 rue Henri Huchard, 75018 Paris, France. ¹⁰Faculté de médecine and Université Denis Diderot Paris VII, 5 Rue Thomas Mann, 75013 Paris, France. ¹¹Centre national de la recherche scientifique (CNRS), unité mixte de recherche 8113, Ecole Normale Supérieure de Cachan, 61 avenue du président Wilson, 94230 Cachan, France. ¹²Laboratoire d'Excellence en Recherche sur le Médicament et l'Innovation Thérapeutique (LERMIT), Campus Paris Saclay, 5 rue Jean-Baptiste Clément 92296 Châtenay-Malabry, France. ¹³Women's Hospital and Harvard Medical School, 25 Shattuck St, Boston, Massachusetts 02115, USA.

¹⁴INSERM, unité mixte de recherche 779, Université Paris XI, Le Kremlin-Bicêtre, France. ¹⁵Institut Cochin, INSERM, unité mixte de recherche 1016, centre national de la recherche scientifique (CNRS), unité mixte de recherche 8104, Université Paris Descartes, et Assistance publique – Hôpitaux de Paris, Hôpitaux Universitaires Paris Centre, Hôpital Cochin, Service d'hématologie biologique, 27 rue du Faubourg Saitn-Jacques, 75014 Paris, France. ¹⁶Centre anticancéreux George François Leclerc, 1 rue professeur Marion, 21079 Dijon, France. ¹⁷Service d'hématologie, Faculté de médecine Paris Descartes, Sorbonne Paris-Cité et Assistance publique – Hôpitaux de Paris Hôpital Necker, 149 rue de Sèvres, 75015 Paris, France.

*These authors contributed equally to this work. §These authors jointly supervised this work.

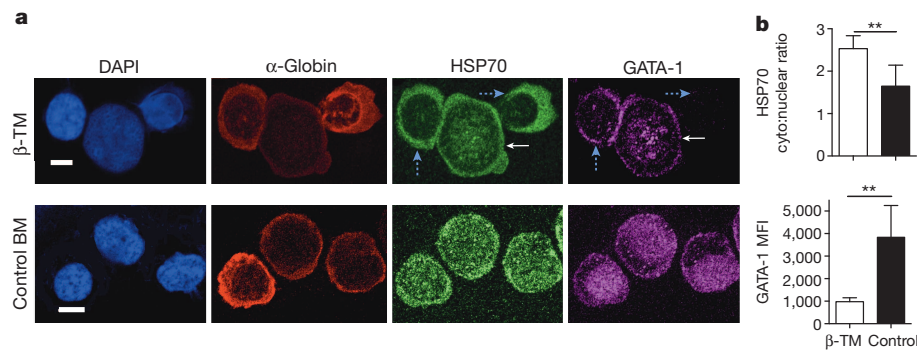


Figure 1 | HSP70 and GATA-1 expression in fresh bone marrow from β -TM patients.

a, Representative confocal microscopy analysis of α -globin, HSP70 and GATA-1 expression in bone marrow (BM) erythroblasts from three β -TM patients and three healthy donors (control). Scale bars, 5 μ m. Blue arrows indicate HSP70 cytoplasmic sequestration in mature β -TM erythroblasts; white arrow, immature erythroblast. **b**, HSP70 cytoplasmic:nuclear MFI ratio (top) and GATA-1 nuclear MFI (bottom) (6–10 cells per experiment). Data are presented as median; error bars are IQR. *P* values determined using the Mann–Whitney *U*-test. *******P* < 0.01.

of these proteins during differentiation. In β -TM-derived erythroblasts, nuclear HSP70 and GATA-1 expression decreased with erythroid differentiation, while it increased in control erythroblasts (Fig. 2d). To ensure that these findings were not due to modifications of the kinetics of differentiation, at day 8 we performed similar experiments from cytoplasmic and nuclear fractionations of sorted erythroblasts according to their band-3 and α_4 -integrin expression¹⁰ (Extended Data Fig. 2a,b). HSP70 and GATA-1 nuclear contents in β -TM polychromatophilic

cells were decreased as compared to less mature β -TM erythroblasts, whereas this was not the case in controls (Extended Data Fig. 2c). In addition, maturation arrest in β -TM patients assessed by flow cytometry was confirmed since TMI was 23.5% (IQR 17–24%) in β -TM cells compared to 39% in controls (IQR 31.6–42%) (*P* = 0.01).

To demonstrate that HSP70 could act as a molecular chaperone of free α -globin chains, we first analysed the subcellular localization of both proteins by co-immunofluorescence experiments (*n* = 7). From

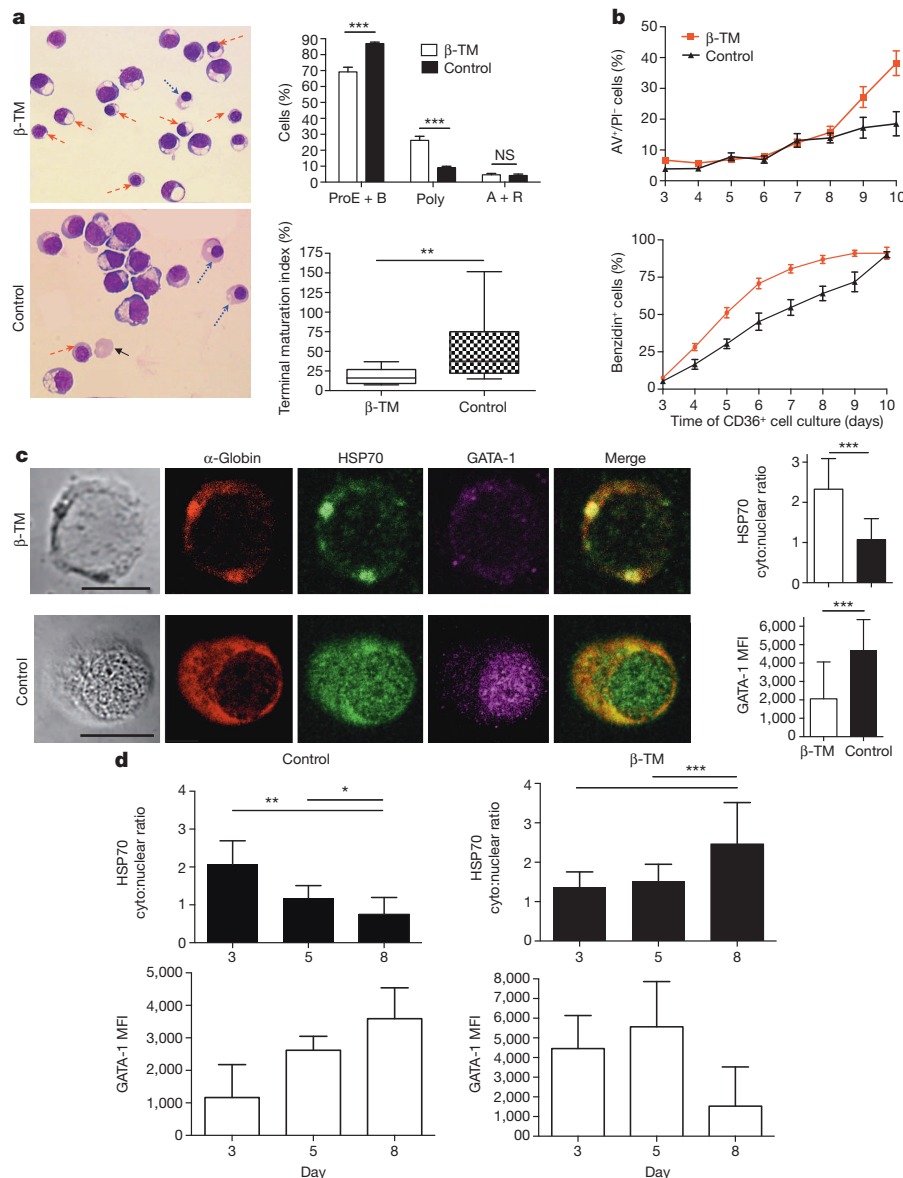


Figure 2 | Characteristics of β -TM ineffective erythropoiesis and kinetics of HSP70 and GATA-1 expression during *in vitro* erythroid differentiation.

CD36⁺ cells derived from CD34⁺ adult β -TM or healthy (control) peripheral blood cells were cultured as described in Methods. **a**, May–Grünwald–Giemsa (MGG) staining at day 8 of CD36⁺ cell culture. Left panels show a representative morphological analysis ($\times 40$ magnification) of erythroid differentiation. Solid black arrows indicate reticulocytes, dotted blue arrows indicate acidophilic cells and dashed red arrows indicate polychromatophilic cells. Upper-right graph represents proportion of each developmental stage. Data are presented as mean \pm standard error of the mean (s.e.m.). ProE + B, proerythroblasts and basophilic cells; Poly, polychromatic cells; A + R, acidophilic cells and reticulocytes. Lower-right graph represents terminal maturation index, as defined in Methods; box plots whiskers represent minimum and maximum values. β -TM group, *n* = 14 independent experiments; control group, *n* = 8. **b**, Top panel: apoptosis curves analysis. Daily mean percentage \pm s.e.m. of Annexin V (AV)-positive, propidium iodide (PI)-negative cells assessed by flow cytometry. Bottom panel: differentiation curves assessed by benzidine staining (mean percentage \pm s.e.m. of positive cells). β -TM group, *n* = 17 independent experiments; control group, *n* = 9. **c**, Left: representative confocal microscopy analysis of HSP70, GATA-1 and α -globin expression at day 8. Scale bar, 5 μ m. Right: HSP70 cytoplasmic:nuclear MFI ratio and nuclear GATA-1 MFI. Data are presented as median (error bars are IQR) for 7 to 31 cells from 10 independent experiments in the β -TM group (from 7 different patients), and 6 in the control group. **d**, Kinetics of HSP70 and GATA-1 expression *in vitro*. HSP70 cytoplasmic:nuclear MFI ratio and GATA-1 nuclear MFI were analysed at days 3, 5 and 8 (8 to 23 cells were analysed each day for each experiment, *n* = 3 independent experiments). Data are presented as median; error bars are IQR. *P* values determined by Mann–Whitney *U*-test. ******P* < 0.05; *******P* < 0.01; ********P* < 0.001; NS, not significant.

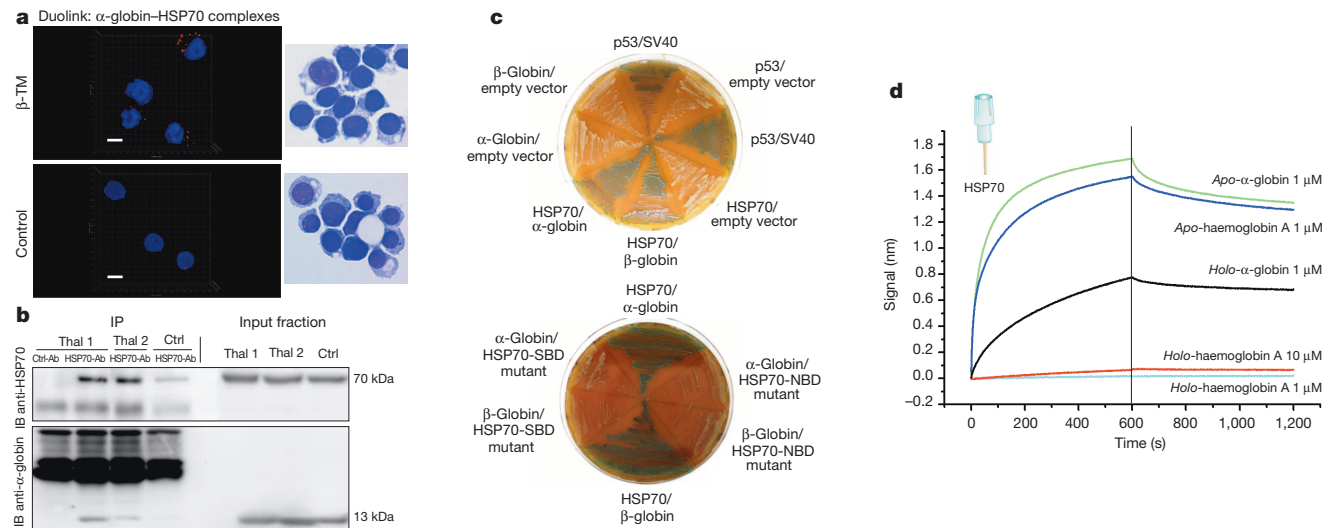


Figure 3 | HSP70 and α -globin interaction. **a**, Left panels: Duolink proximity ligation assay using anti-HSP70 and anti- α -globin antibodies in polychromatophilic erythroblasts purified by flow cytometry at day 8 of CD36⁺ cell culture. Right panels: morphological analysis of purified cells (MGG). The red spots indicate close proximity (<40 nm) between bound antibodies. Nuclei are stained with DAPI (blue). Scale bar, 5 μ m. **b**, Immunoprecipitation (IP)-immunoblot (IB) analysis of whole cell lysates from day 8 cultured erythroblasts derived from two β -TM patients (Thal 1 and 2) and one healthy adult (Ctrl, control). Immunoprecipitation with anti-HSP70 monoclonal antibody (HSP70-Ab) or control antibody (Ctrl-Ab) was followed by immunoblotting using an anti- α -globin antibody (bottom panel).

day 6 of culture, HSP70 and α -globin co-localized in the cytoplasm of β -TM erythroblasts. Features of aggregates were sometimes observed in differentiated, haemoglobinized cells (representative image in Fig. 2c). Co-localization was quantified using the average Pearson's correlation coefficient (r). At day 8, an apparent high level of co-localization between HSP70 and α -globin was detected, both in β -TM ($r = 0.4 \pm 0.13$) and in controls ($r = 0.31 \pm 0.09$). Similar findings were observed in bone marrow experiments from β -TM patients and healthy donors ($r = 0.46 \pm 0.14$ and 0.36 ± 0.08 , respectively). To characterize this co-localization further, we used an *in situ* close proximity ligation assay (Duolink), which allows the identification of interacting proteins as fluorescent spots. At day 8, the number of spots per cell in β -TM erythroblasts was dependent on the degree of differentiation and thus on the accumulation of α -globin chains (Extended Data Fig. 3). Likewise, more spots were observed in β -TM purified polychromatophilic erythroblasts compared to controls (Fig. 3a).

Interaction between HSP70 and α -globin was then demonstrated by co-immunoprecipitation experiments in whole erythroblast lysates at day 8. This approach showed that soluble α -globin chains interacted with HSP70 (Fig. 3b). Next, using a yeast two-hybrid system, we provided additional evidences for this interaction. Blue diploid transformants could be detected on a high stringency minimal medium, indicating a direct HSP70- α -globin biochemical interaction. Similar results were obtained when the β -globin coding sequence was used as prey (Fig. 3c). To identify HSP70 domains involved in this interaction, we tested the binding of α -globin chains to deletion mutants that expressed either the nucleotide binding domain or the substrate binding domain of HSP70 (Fig. 3c). Notably, neither of these two mutants interacted with α -globin chains, suggesting that the entire structure of HSP70 is required for the recognition of α -globin.

To improve characterization of HSP70- α -globin complexes, we performed protein-protein interaction experiments using biolayer interferometry. We found that HSP70 binds to apo- α -globin chains. Interestingly, the binding affinity of HSP70 for holo- α -globin chains was strongly decreased and no interaction was observed between HSP70 and holo-haemoglobin-A (Fig. 3d).

Immunoblotting with anti-HSP70 polyclonal antibody (top panel) was performed as control. Input fractions (3% of the starting material, right panels) were included as controls. **c**, Top panel: yeast two-hybrid assay detected a direct interaction of α - and β -globin chains with HSP70 (blue diploid yeast cells). This interaction was lost with HSP70 deletion mutants. SBD, substrate binding domain; NBD, nucleotide binding domain. **d**, The association between purified HSP70 and α -globin chains (apo and holo, 1 μ M) as well as holo-haemoglobin A (1 and 10 μ M) and apo-haemoglobin A (1 μ M) was analysed using biolayer interferometry. Data in panels **a**, **c** and **d** are representative of three independent experiments.

α -Haemoglobin stabilizing protein (AHSP) is a well reported cognate chaperone of α -globin chains¹¹. Therefore, we compared the dissociation constant (K_d) of apo- α -globin for both proteins, HSP70 and AHSP. We found, using biolayer interferometry, that both chaperones bind apo- α -globin with a comparable K_d (4.65 nM and 8.35 nM for HSP70 and AHSP, respectively) (Extended Data Fig. 4).

These results show that HSP70 may, like AHSP, act as a chaperone of α -globin chains. Kinetics studies of cytoplasmic HSP70 and AHSP expression during erythropoiesis demonstrated that while AHSP expression was relatively stable during differentiation and maturation, HSP70 expression increased with haemoglobinization and erythroid differentiation (Extended Data Fig. 5). This finding suggests that the increase in HSP70 expression might be related to cellular stress⁹ induced by high synthesis of globin chains and that HSP70 could substitute for AHSP chaperone function at the later stages of erythroid differentiation. However, this cytoprotective function of HSP70 might be detrimental in the case of globin chain imbalance in β -TM by preventing HSP70 nuclear localization.

To investigate the contribution of HSP70 cytoplasmic sequestration to the pathophysiology of β -TM ineffective erythropoiesis, we used lentiviral transduction to restore HSP70 expression in the nucleus of β -TM erythroblasts. For this purpose, β -TM CD34⁺ cells were transduced with lentiviruses expressing a nuclear-targeted HSP70 mutant (HSP70 (S400A))¹², wild-type HSP70 or an empty lentivector. As expected, at day 7 of differentiation, HSP70(S400A) (Fig. 4a) and wild-type HSP70 (not shown) lentivectors increased nuclear HSP70 localization and rescued GATA-1 expression in β -TM erythroblasts. Restoration of nuclear HSP70 localization efficiently improved the terminal maturation of β -TM erythroblasts. At day 7, in HSP70(S400A) transduced β -TM erythroblasts, the percentage of mature cells (acidophilic cells and reticulocytes) was increased when compared to the empty control vector ($10.6 \pm 3.3\%$ versus $1.1 \pm 0.7\%$; $P = 0.01$) (Fig. 4b). Similarly, the TMI was increased both by cytological and flow cytometry analysis: $31 \pm 13.4\%$ versus $4.3 \pm 3.3\%$ ($P = 0.01$) and $41.3 \pm 8.2\%$ versus $18.2 \pm 4.7\%$ ($P = 0.01$) (Fig. 4c), respectively. In addition, rescuing nuclear HSP70 localization induced a twofold decrease in apoptosis ($9.9 \pm 2.0\%$ versus 20.7

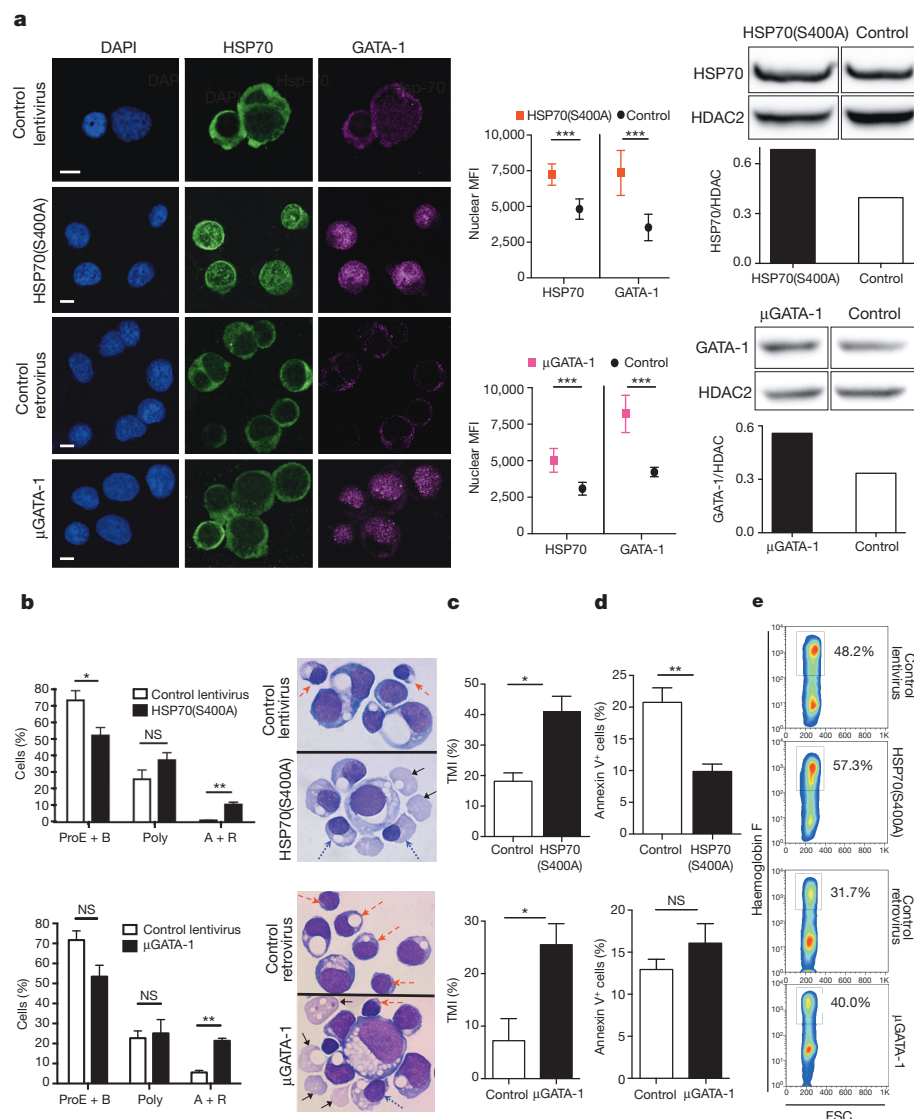


Figure 4 | Transduction of HSP70(S400A) rescues erythroid terminal maturation and cell survival, and μ GATA-1 rescues terminal maturation of β -TM erythroblasts. β -TM CD34⁺ cells were transduced with a nuclear-targeted HSP70 lentiviral mutant (HSP70(S400A)), a retroviral mutant of GATA-1 uncleavable by activated caspase-3 (μ GATA-1), and their appropriate empty vectors as control and cultured as described above. All data were analysed at day 7 of CD36⁺ cell culture. **a**, A confocal microscopy image of HSP70 and GATA-1 is shown in the left panel (representative of three experiments). Graphs in the middle panel represent the mean nuclear MFI of HSP70 and GATA-1 (\pm 95% confidence intervals) in transduced and control β -TM erythroblasts (10 to 20 cells per field analysed, $n = 3$ independent experiments). Scale bars, 5 μ m. Western blot analyses of nuclear HSP70 and GATA-1 from transduced cells are shown in the right panel. Graphs show optical density values of specific protein normalized to that of histone deacetylase 2 (HDAC2, loading control).

b, Erythroid differentiation. Graphs represent the proportion of each erythroid stage for each transduction after MGG staining. One representative morphological analysis is shown ($\times 100$ magnification). Solid black arrows indicate reticulocytes, dotted blue arrows indicate acidophilic cells and dashed red arrows indicate polychromatophilic cells. ProE + B, proerythroblasts and basophilic cells; Poly, polychromatic cells; A + R, acidophilic cells and reticulocytes. **c**, TMI calculated by flow cytometry after α_4 -integrin/band-3 staining is shown. **d**, Apoptosis in the green fluorescent protein (GFP)⁺/propidium iodide⁻ cell population was assessed by Annexin V binding by flow cytometry. **e**, The percentage of cells containing a high level of fetal haemoglobin was assessed by flow cytometry on mature cells (GFP⁺ and low forward light scatter (FSC)). A representative experiment of each transduction is shown ($n = 3$). Data in panels **b–d** are presented as mean \pm s.e.m. for three independent experiments for each transduction. P values in panels **a–d** were determined by two-tailed Student's unpaired t -tests. * $P < 0.05$; ** $P < 0.01$; NS, not significant.

$\pm 3.9\%$; $P = 0.01$) (Fig. 4d), but no change in cell proliferation (data not shown). As a result the absolute number of very mature erythroid cells was significantly increased (Extended Data Fig. 6).

For further confirmation that decrease in maturation and apoptosis was due to the lack of nuclear protective effect of HSP70 on GATA-1, we transduced β -TM CD34⁺ cells with a GATA-1 mutant that was uncleavable by caspase-3 (μ GATA-1)¹³ or an empty vector. The μ GATA-1 mutant had a positive effect on erythroid terminal maturation that was similar to that of HSP70(S400A) (Fig. 4a–c). However, apoptosis was not corrected (Fig. 4d), indicating that cleavage of GATA-1 mainly impairs erythroid maturation but contributes to a lesser extent to apoptosis of β -TM cells, as previously reported in the ineffective erythropoiesis of low-grade myelodysplastic syndromes¹².

Fetal haemoglobin (HbF, comprising two α and two γ subunits), which is replaced after birth by adult haemoglobin (made up of two α and two β subunits), is concentrated in a few F cells and represents less than 1% of the haemoglobin content in healthy adults¹⁴. In β -TM patients, there is an increased proportion of F cells to compensate for the lack of β -chain synthesis, and the only surviving mature erythroblasts are F cells. GATA-1 has a major role in regulating haemoglobin gene expression; it is sometimes described as a transcription repressor or activator of the human γ -globin chain gene^{15–18}. Thus, we studied the effect of GATA-1

nuclear restoration on HbF expression by flow cytometry. At day 7, we observed that the percentage of cells containing a high level of fetal haemoglobin was significantly increased concomitantly with the protection of GATA-1 by HSP70(S400A) ($54.8 \pm 12\%$ versus $45.9 \pm 10.5\%$; $P < 0.004$) and also by μ GATA-1 ($51.4 \pm 8.2\%$ versus $40.5 \pm 9.4\%$; $P < 0.002$) (Fig. 4e). Finally, to ensure the specificity of our findings, we transduced β -TM CD34⁺ cells with the β -globin gene using lentiviral technology and found that this led to HSP70 nuclear re-localization, GATA-1 protection and restoration of normal erythroid maturation (Extended Data Fig. 7).

Taken together, our data demonstrate that HSP70 cytoplasmic sequestration by an excess of α -globin chains prevents its nuclear localization and thus its protective function against GATA-1 cleavage. Blockade of end-stage maturation of polychromatophilic erythroblasts is one of the key mechanisms that induce β -TM ineffective erythropoiesis. It is tempting to hypothesize that besides its function as chaperone of GATA-1, HSP70 might have been selected during evolution to serve as a complementary chaperone of AHSP to prevent an excess of free globin chains at later stages of erythroid differentiation.

In addition to ineffective erythropoiesis, β -TM dyserythropoiesis is characterized by an abnormal *in vivo* expansion of immature erythroid progenitors^{5,6,19} due to an increase of erythropoietin production and

activation of the erythropoietin–Janus-kinase-2 pathway¹⁹, which acts synergistically with proliferative signals provided by bone-marrow macrophages in erythroblastic islands²⁰. Moreover, studies have recently shown that in β -TM, erythroid progenitors and macrophages produce a high level of growth differentiation factor (GDF)-11 (refs 21, 22). GDF-11 is a protein of the transforming-growth-factor- β family, which contributes to erythroid cell expansion by reducing Fas/FasL expression and maturation arrest in mouse models of β -thalassaemia by a poorly characterized mechanism^{21,22}. On the basis of these works, it has been shown in mouse models of β -thalassaemia intermedia that inhibition of Jak2 (ref. 19) or GDF-11 (refs 21, 22), or macrophage depletion²⁰, may be possible therapeutic strategies.

Although our *in vitro* human model cannot reconstitute all the features of β -TM dyserythropoiesis *in vivo* (for example, lack of bone marrow microenvironment including macrophages), our findings provide a rationale for a new targeted therapy to improve ineffective erythropoiesis. Therefore, small compounds disrupting the HSP70– α -globin complex in the cytoplasm or decreasing HSP70 nuclear export may increase the nuclear localization of HSP70 and thus GATA-1 protection. This may in turn reduce maturation arrest and increase the number of fetal cells. Ultimately, these outcomes may decrease hepcidin repression and patients' requirement for blood transfusions, thereby reducing life-threatening iron overload.

METHODS SUMMARY

Erythroblasts were generated *in vitro* from peripheral blood circulating CD34⁺ cells from seven adults with β^0 -TM (thalassaemia major without any residual production of β chains) or eight healthy donors as previously described⁸. Fresh normal bone marrow smears were obtained from three adult patients with β^0 -TM who had undergone cholecystectomy or splenectomy, or three healthy controls (allogenic bone marrow donors), after they had given written informed consent. This study was performed according to the Helsinki Declaration with approval from the ethics committee of our institution. Details of reagents and protocols for cell proliferation and differentiation analysis, confocal fluorescence microscopy analysis, yeast two-hybrid assays, biolayer interferometry (Octet), production of lentiviral particles and transduction of haematopoietic progenitors, and generation of HSP70 and GATA-1 mutants are provided in the online Methods. Statistical analyses were performed using GraphPad Prism software <http://www.graphpad.com/scientific-software/prism/>. Data are expressed as mean \pm standard deviation (s.d.) or median with the IQR in parentheses, unless noted otherwise. Two-tailed Student's unpaired *t*-test or Mann–Whitney *U*-test was used as appropriate. *P* values <0.05 were considered significant; values smaller than this are indicated in figure legends.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 February 2013; accepted 25 June 2014.

Published online 24 August 2014.

- Khandros, E. & Weiss, M. J. Protein quality control during erythropoiesis and hemoglobin synthesis. *Hematol. Oncol. Clin. North Am.* **24**, 1071–1088 (2010).
- Ginzburg, Y. & Rivella, S. β -thalassaemia: a model for elucidating the dynamic regulation of ineffective erythropoiesis and iron metabolism. *Blood* **118**, 4321–4330 (2011).
- Yuan, J. *et al.* Accelerated programmed cell death (apoptosis) in erythroid precursors of patients with severe β -thalassaemia (Cooley's anemia). *Blood* **82**, 374–377 (1993).
- Mathias, L. A. *et al.* Ineffective erythropoiesis in β -thalassaemia major is due to apoptosis at the polychromatophilic normoblast stage. *Exp. Hematol.* **28**, 1343–1353 (2000).
- Centis, F. *et al.* The importance of erythroid expansion in determining the extent of apoptosis in erythroid precursors in patients with β -thalassaemia major. *Blood* **96**, 3624–3629 (2000).
- Ribeil, J. A. *et al.* Ineffective erythropoiesis in β -thalassaemia. *ScientificWorldJournal* **2013**, 394295 (2013).
- Zermati, Y. *et al.* Caspase activation is required for terminal erythroid differentiation. *J. Exp. Med.* **193**, 247–254 (2001).

- Ribeil, J.-A. *et al.* Hsp70 regulates erythropoiesis by preventing caspase-3-mediated cleavage of GATA-1. *Nature* **445**, 102–105 (2007).
- Hartl, F. U., Bracher, A. & Hayer-Hartl, M. Molecular chaperones in protein folding and proteostasis. *Nature* **475**, 324–332 (2011).
- Hu, J. *et al.* Isolation and functional characterization of human erythroblasts at distinct stages: implications for understanding of normal and disordered erythropoiesis *in vivo*. *Blood* **121**, 3246–3253 (2013).
- Kihm, A. J. *et al.* An abundant erythroid protein that stabilizes free α -haemoglobin. *Nature* **417**, 758–763 (2002).
- Frisan, E. *et al.* Defective nuclear localization of Hsp70 is associated with dyserythropoiesis and GATA-1 cleavage in myelodysplastic syndromes. *Blood* **119**, 1532–1542 (2012).
- De Maria, R. *et al.* Negative regulation of erythropoiesis by caspase-mediated cleavage of GATA-1. *Nature* **401**, 489–493 (1999).
- Dover, G. J. & Boyer, S. H. Fetal hemoglobin-containing cells have the same mean corpuscular hemoglobin as cells without fetal hemoglobin: a reciprocal relationship between gamma- and beta-globin gene expression in normal subjects and in those with high fetal hemoglobin production. *Blood* **69**, 1109–1113 (1987).
- Yao, X. *et al.* Role of STAT3 and GATA-1 interactions in gamma-globin gene expression. *Exp. Hematol.* **37**, 889–900 (2009).
- Woon Kim, Y., Kim, S., Geun Kim, C. & Kim, A. The distinctive roles of erythroid specific activator GATA-1 and NF-E2 in transcription of the human fetal γ -globin genes. *Nucleic Acids Res.* **39**, 6944–6955 (2011).
- Zhu, J. *et al.* Recombinant erythroid Kruppel-like factor fused to GATA1 up-regulates δ - and γ -globin expression in erythroid cells. *Blood* **117**, 3045–3052 (2011).
- Sankaran, V. G. & Orkin, S. H. The switch from fetal to adult hemoglobin. *Cold Spring Harb. Perspect. Med.* **3**, a011643 (2013).
- Libani, I. V. *et al.* Decreased differentiation of erythroid cells exacerbates ineffective erythropoiesis in beta-thalassemia. *Blood* **112**, 875–885 (2008).
- Ramos, P. *et al.* Macrophages support pathological erythropoiesis in polycythemia vera and β -thalassaemia. *Nature Med.* **19**, 437–445 (2013).
- Dussiot, M. *et al.* An activin receptor IIA ligand trap corrects ineffective erythropoiesis in β -thalassaemia. *Nature Med.* **20**, 398–407 (2014).
- Suragani, R. N. V. S. *et al.* Modified activin receptor IIB ligand trap mitigates ineffective erythropoiesis and disease complications in murine β -thalassaemia. *Blood* <http://dx.doi.org/10.1182/blood-2013-06-511238> (19 June 2014).

Acknowledgements J.-B.A. was a recipient of grants from the Société Nationale Française de Médecine Interne (SNFMI), Société Française d'Hématologie (SFH), Centre National de la Recherche Scientifique (CNRS) and Assistance publique – Hôpitaux de Paris (AP-HP) (poste d'accueil APHP-CNRS), and from Groupe Pasteur Mutualité. S.D. was a recipient of funding from Institut FARMAN (ENS de Cachan). I.C.B. was a recipient of a PhD scholarship fund from ENS de Cachan. A.H. was a recipient of a post-doctoral grant from the Conseil Régional de Bourgogne and M.S. from l'Institut National du Cancer, INCa. This program has received a state subsidy managed by the National Research Agency under the 'Investments for the Future' program bearing the reference ANR-10-IAHU-01, Ligue nationale contre le cancer, Fondation pour la recherche médicale (FRM) and Fondation de France. The labex GR-Ex is funded by the program 'Investments for the Future' of the French National Research Agency, reference ANR-11-IDEX-0005-02. We would also like to thank N. Goudin and R. Desvaux from the Plateforme d'imagerie cellulaire de l'IFR94, Faculté de Médecine Necker (Paris, France) for experimental help in image acquisition and analysis on confocal microscopy, and the Department of Biotherapy at the Necker Hospital (Paris, France) for providing blood samples. We thank P. England from CNRS 1129, unité de biochimie cellulaire, Institut Pasteur, Paris, for his experimental help. We thank E. Frisan from the Department of Haematology and INSERM U1016, Hôpital Cochin, Paris, and A. de Thonel from INSERM U866, Dijon, for the gift of HSP70(S400A) mutant. We thank U. Testa, A. Zeuner and R. de Maria from the Department of Haematology and Oncology, Instituto Superiore di Santia, Rome, Italy, for the gift of cDNAs of GATA-1. We also thank C. Brouzes from the Department of biological haematology for her help for the smears' reading.

Author Contributions J.-B.A. designed and performed all experiments, analysed the data and wrote the manuscript. J.-A.R. designed the study, supervised the overall project, provided human samples, analysed the data and helped to write the manuscript. A.H., M.S. and C.G. performed the yeast two-hybrid assays and analysed the data. G.M. and R.S. performed the biolayer interferometry (Octet) and helped to write the manuscript. C.G., S.D., I.C.B., L.T., C.A., V.B.-C. and Y.B. performed experiments to characterise the HSP70/ α -globin complex and helped to write the manuscript. F.G. performed biochemical experiments and helped to write the manuscript. M.D., I.C.M., Z.B.-C., M.F. and T.T.M. analysed the data and helped to write the manuscript. O.N., P.L., S.C. and Y.B. performed the β -globin gene lentiviral transduction and HPLC analyses. O.H. designed the study, supervised the overall project, analysed the data and wrote the manuscript. G.C. performed experiments, supervised the overall project, analysed the data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to O.H. (ohermine@gmail.com) or G.C. (genevieve_courtois@yahoo.fr).

METHODS

Erythroid liquid culture. The circulating peripheral blood from β -TM patients contains a small number of haematopoietic progenitor cells²³. Erythroid cells were generated *in vitro* from circulating peripheral blood CD34⁺ cells from adult patients with β^0 -thalassaemia major (TM without any residual production of β chains), which were collected before routine transfusion, or from control patients who were healthy donors treated with granulocyte colony-stimulating factor to induce haematopoietic stem cell mobilization.

This study was done according to the Helsinki Declaration with the approval from the ethics committee of our institution (Comité de Protection des personnes Ile de France II). All patients gave written informed consent. In the first step of culture ('cell expansion'), isolated CD34⁺ progenitors (Miltenyi CD34 Progenitor Cell Isolation Kit) were grown in the presence of 100 ng ml⁻¹ IL-6, 10 ng ml⁻¹ IL-3 and 100 ng ml⁻¹ SCF for 7 days. At day 7, CD36⁺ erythroid progenitors were isolated by magnetic isolation (MiltenyiBiotec). In the second phase of culture, which allows the differentiation and maturation of erythroblasts, CD36⁺ cells were cultured in the presence of 10 ng ml⁻¹ IL-3, 50 ng ml⁻¹ SCF and 2 U ml⁻¹ erythropoietin in IMDM (Gibco cell culture) supplemented with 15% BIT 9500 (Stem Cell Technologies), as previously described^{8,24,25}. Day 0 of differentiation phase is the start of CD36⁺ erythroblasts culture.

All samples were analysed except those with aberrant mortality at day 6 of the CD36⁺ cell culture (>40% cell death). This exclusion criterion was pre-specified and required to avoid any technical problem in this culture system.

Apoptosis assay and cell differentiation. Apoptosis was assessed by Annexin V binding and propidium iodide staining (e-bioscience). Early apoptotic cells were defined as Annexin-V-positive and propidium-iodide-negative.

Differentiation was assessed by various methods. First, morphological analysis after May-Grünwald-Giemsa (MGG) staining was used. Cells were examined under a Leica DMRB microscope with a PLFluotar $\times 40$ oil objective in a blinded fashion. The number of proerythroblasts, basophilic, polychromatic and acidophilic cells, and reticulocytes was assessed in each experiment by counting approximately 300 cells in consecutive oil immersion fields and expressed as a percentage of total cells. Additionally, differentiation was assessed by calculating a terminal maturation index on the basis of MGG staining, defined by the number of (acidophils + reticulocytes per slide) $\times 100$ divided by the number of polychromatophilic cells per slide. This allowed us to better characterize the maturation arrest at the polychromatophilic stage, which is known to be a hallmark of β -TM ineffective erythropoiesis⁴, and its modulation. We also measured the haemoglobin content, as assessed by benzidine staining. Flow cytometry analysis was also performed at several points of erythroid culture after double labelling with a phycoerythrin-conjugated anti-glycophorin-A (GpA, BD Pharmingen) antibody and an allophycocyanin (APC)-conjugated anti-CD117 (also known as c-Kit) (e-bioscience) antibody. The percentage of GpA⁺/CD117⁺ cells represented mature erythroid cells²⁵. Band 3 and α_4 -integrin were recently described as optimal surface markers for developmental stages of erythropoiesis, particularly to differentiate highly mature erythroblasts¹⁰. Phycoerythrin-conjugated anti-band-3 (PE-BRIC6 conjugate, Bristol Institute for Transfusion Sciences) and APC-conjugated anti- α_4 -integrin (MiltenyiBiotec) antibodies were used for flow cytometry. This double labelling allowed the assessment of terminal erythroblastic differentiation and purification cell populations by cell sorting (BD FACSaria II SORP). Gates I (band 3^{low}; α_4 -integrin^{high}), II (band 3^{med}; α_4 -integrin^{low}) and III (band 3^{high}; α_4 -integrin^{low}) were defined and cells from each gate were sorted and analysed morphologically after MGG staining. The majority (around 80%) of cells at each gate were: gate I, basophilic erythroblasts; gate II, polychromatophilic erythroblasts; and gate III, acidophilic erythroblasts (Extended Data Fig. 2). Terminal maturation index by flow cytometry was defined as (percentage of cells in gate III) $\times 100$ /(percentage of cells in gate II).

F cells, containing a high level of fetal haemoglobin, were evidenced by flow cytometry analysis. The cultured cells were fixed and permeabilized, washed with 1% BSA in 1X PBS and then stained with phycoerythrin-conjugated anti-human-fetal-haemoglobin (BD Pharmingen) for 30 min at room temperature. The cells were then analysed with a FACSCalibur (Becton Dickinson).

Analysis of the HSP70- α -globin complex. *Yeast two-hybrid assay.* The bait vector used was pGBKT7 and the prey vector was pGADT7 (Clontech). The human α - and β -globin coding sequences were cloned into the EcoRI/BamHI and NdeI/ClaI sites, respectively, of the pGADT7 vector. The coding sequence of the human HSP70 was cloned into the EcoRI/BamHI sites of the pGBKT7 vector. An N-terminal nucleotide-binding domain (residues 1–380) and a C-terminal substrate-binding domain (residues 394–615) were cloned in pGBKT7. The pGBKT7-p53 plasmid was used as a positive bait control. This plasmid encodes the GAL4 DNA-binding domain fused with murine p53. As a positive control prey plasmid, pGADT7-SV40, which encodes the GAL4 activation domain fused with SV40 large T antigen, was used. The empty vector, pGADT7, was used as a negative control prey plasmid. All pGADT7- and pGBKT7-derived vectors were transformed into Y187 and Y2HGOLD

yeast strains, respectively (Clontech Yeastmaker Yeast Transformation System 2). After growth at 30 °C for 3 to 5 days, the transformants were selected on Leu₋ and Trp₋ minimal media plates, respectively. Each prey strain was mated with the bait strain to generate diploid yeast cells (Clontech Matchmaker Gold Yeast 2-Hybrid system). Diploid yeast cells selected on Leu₋ Trp₋ minimal media plates were then patched onto Leu₋ Trp₋ minimal media plates with X- α -Galactosidase (40 μ g ml⁻¹) and Aureobasidin A (70 ng ml⁻¹). Blue diploid cells appeared after 3 to 5 days at 30 °C, indicating the interaction between the bait and the prey proteins. To confirm these results, the diploid yeast cells were then patched onto higher stringency His₋ Ade₋ Leu₋ Trp₋ minimal media plates supplemented with 40 μ g ml⁻¹ X- α -Galactosidase and 70 ng ml⁻¹ Aureobasidin A.

Confocal analysis. The cells were washed, spun onto slides, fixed with acetone, hydrated with cold 1% BSA in 1X PBS for 30 min, treated with formaldehyde (Sigma) for 15 min, and then with methanol (Prolabo) for 10 min at room temperature. Next, the cells were permeabilized with 0.2% Triton X100 (Sigma) in 1X PBS for 10 min at 4 °C, washed with 1% BSA in 1X PBS and incubated in 3% BSA for 30 min. They were then sequentially incubated with the antibodies as follows: anti-GATA-1 overnight at 4 °C (GATA-1 N6, SC-265 Santa Cruz), anti-rat-Cy3 (Jackson) for 45 min at room temperature, rabbit anti-HSP70 (SPA 812, Enzo Lifesciences) for 1 h at room temperature, anti-rabbit Alexa 647 for 45 min at room temperature and anti- α -haemoglobin-FITC (Abcam) for 60 min at room temperature. All antibodies were diluted in 1% BSA and 0.1% Tween (Sigma) in 1X PBS. Nuclei were stained with DAPI, and the slides were examined with a confocal laser microscope (LSM 510 Carl Zeiss). Image analysis was performed using ImageJ software. For each independent experiment, 3–6 fields were collected per slide. Each field was selected by viewing α -globin staining and cell size to identify the highest number of the most differentiated cells in each field (small cells and higher α -globin intensity). For each field, DAPI-positive staining was used to define the nuclear region of interest (ROI) and α -globin staining to define the cytoplasmic ROI. Bright-field images were also used to demarcate each cellular compartment and detect viable cells. Each of these ROI masks was then applied, by the image calculator, to obtain quantitative MFI data. About 20 cells were analysed per experiment. The number of cells analysed and the number of independent experiments is specified in the figure legends. Fresh, healthy bone-marrow cell smears were fixed with acetone. Permeabilization and labelling with anti-GATA-1, anti-HSP70 and anti- α -globin antibodies were performed as above. Co-localization was assessed and quantified using the average Pearson's correlation coefficient with the JACoP plugin, using ImageJ software. It was confirmed by Van Steensel's approach as already described²⁶.

To analyse the HSP70- α -globin interaction more precisely, we used the Duolink II technology (Olink Bioscience), which is an *in situ* proximity ligation assay technology. In this assay, a pair of secondary antibodies labelled with oligonucleotides (PLA probes) only generates a signal when the two probes are bound in close proximity (<40 nm). The signal from each detected pair of PLA probes is visualized as an individual fluorescent spot²⁷. Slides were incubated with primary antibodies as described above and with secondary antibodies conjugated with oligonucleotides (PLA probe MINUS anti-mouse and PLA probe PLUS anti-rabbit). Ligation and amplification reactions were performed according to the manufacturer's instructions. **Cell fractionation and immunoprecipitation.** Separate cytoplasmic and nuclear protein fractions were extracted from erythroid progenitors (day 8 of the CD36⁺ transduced cells or from purified erythroblasts at different developmental stages) using NE-PER nuclear and cytoplasmic extraction reagents (Thermo Scientific), following manufacturer's protocol. Protein concentrations were determined using the Pierce BCA protein assay kit (Thermo Scientific) and 10 μ g of nuclear or cytoplasmic protein lysates were resolved on 14% acrylamide gels and analysed by immunoblotting. Antibodies used included rabbit anti-HSP70 SPA 812 (Enzo Lifesciences), mouse anti- α -haemoglobin (Abnova), rabbit anti-AHSP (Rockland), rat anti-Hsc70 1B5 (Stressgen), mouse anti-Hsp90 AC88 (Calbiochem), mouse anti-HDAC2 3F3 (Millipore) and rat anti-GATA1 N1 (Santa Cruz). Antigens were visualized by chemiluminescence using SuperSignal West Dura (Thermo Scientific). Band volumes were analysed using the ImageLab software program and normalized to a HDAC2 blot (nucleus fraction) or a Hsp90 blot (cytoplasmic fraction).

For immunoprecipitation assays, 10⁷ erythroblasts from day 8 of the CD36⁺ cell culture were lysed on ice in an immunoprecipitation buffer (0.4% NP40, 150 mM NaCl, 5 mM EDTA, 65 mM Tris HCl pH 7.5, 50 mM HEPES, 3% glycerol, 1 mM orthovanadate, 1 mM PMSF, 1 mM DTT, and 10 μ g ml⁻¹ of aprotinin, leupeptin and pepstatin) for 5 min, then spun at 14,000 r.p.m. for 2 min and the supernatant was collected. Samples were pre-cleared with protein G Sepharose beads (GE Healthcare, Lifesciences). Whole cell lysates were incubated overnight at 4 °C with mouse anti-HSP70 (1 μ g) monoclonal antibody (SPA 810 Stressgen) or control antibody with gentle rotation. Protein G Sepharose beads were then added and incubated for 45 min. Beads were then washed with PBS containing the immunoprecipitation buffer. Proteins were resolved on 15% acrylamide gels and analysed by immunoblotting using a mouse anti- α -globin monoclonal antibody (Abnova). To ensure

the quality of HSP70 immunoprecipitation, protein lysates (1/10 of the total volume used in 15% acrylamide gel) were also resolved on 8% acrylamide gels and analysed by immunoblotting using a rabbit anti-HSP70 polyclonal antibody (SPA 812 Enzo Lifesciences). Three percent of the starting material, the total cell lysate, before immunoprecipitation (input fraction) was immunoblotted as control.

Biolayer interferometry. Protein–protein interaction experiments were conducted at 25 °C with an OctetRed instrument using purified proteins (ForteBio, Menlo Park, CA, USA). The ligand (either HSP70 or AHSP) was biotinylated using EZ-Link NHS-PEG₄-biotin (Thermo Fisher Scientific) according to the manufacturer's protocol and immobilized at a concentration of 20 µg ml⁻¹ on streptavidin sensors (ForteBio). These functionalized biosensors were dipped into black Greiner 96-well microplates agitated at 1,000 r.p.m. and filled with a solution of 200 µl containing the analyte. Each K_d was determined with a 1:1 stoichiometry model using a global fit with R_{max} unlinked by sensor (ForteBio, Data analysis software version 7.1.0.36). All sensorgrams were corrected for baseline drift by subtracting a control sensor exposed to running buffer only.

The AHSP was expressed as glutathione S transferase fusion protein in BL21(DE3) cells and purified by affinity chromatography on Glutathione Sepharose 4B (GE Healthcare Lifesciences)²⁸. The recombinant AHSP with an N-terminal Gly-Pro-Leu-Gly-Ser peptide was obtained after cleavage by PreScission Protease (GE Healthcare Lifesciences)²⁸.

Haemoglobin A (HbA) was purified from human red blood cells by DEAE-cellulose chromatography performed in 0.2 M glycine buffer, pH 7.8²⁹. The α -subunit chains of HbA (holo- α -chains) were purified by dissociation of HbA in the presence of parahydroxymercuribenzoate followed by ion exchange chromatographies and then saturated with carbon monoxide³⁰. The apo- α -globin chains were purified using the acid-acetone method³¹. This purification method is non-denaturing, thus globin chains used for OctetRed experiments retain their native forms.

Viral transduction. *Lentiviral production.* The nucleus-targeted HSP70 mutant (HSP70(S400A)) was cloned in the pTrip_U3EF1 lentiviral bicistronic vector upstream of an internal ribosome entry site (IRES)-CMV promoter–green-fluorescent-protein (GFP) cassette. Infectious vector particles were produced in 293T cells by cotransfection of the vector with the encapsidation plasmid psPAX2 and the expression plasmid pHCMV-G, using the jetPRIME transfection reagent (Polyplus). Supernatants were collected 48 and 72 h after transfection and were pooled and concentrated by ultracentrifugation. Virus stocks were kept frozen at –80 °C.

For the lentivirus production of the β -globin gene, vesicular stomatitis virus glycoprotein pseudotyped lentiviral supernatant was produced by transient transfection of human embryonic kidney 293T cells with the 5-plasmid system (LentiGlobin construct, HPV 275-gag-pol plasmid, Ψ N 15-vsvG env plasmid, p633-rev plasmid, HPV601-tat plasmid) by calcium phosphate coprecipitation in Dulbecco's modified Eagle's media supplemented with 5% fetal bovine serum (Invitrogen), followed by harvest in CellGro SCGM serum-free media (CellGenix) after 48 h. Concentrated virus was then frozen and stored at –80 °C.

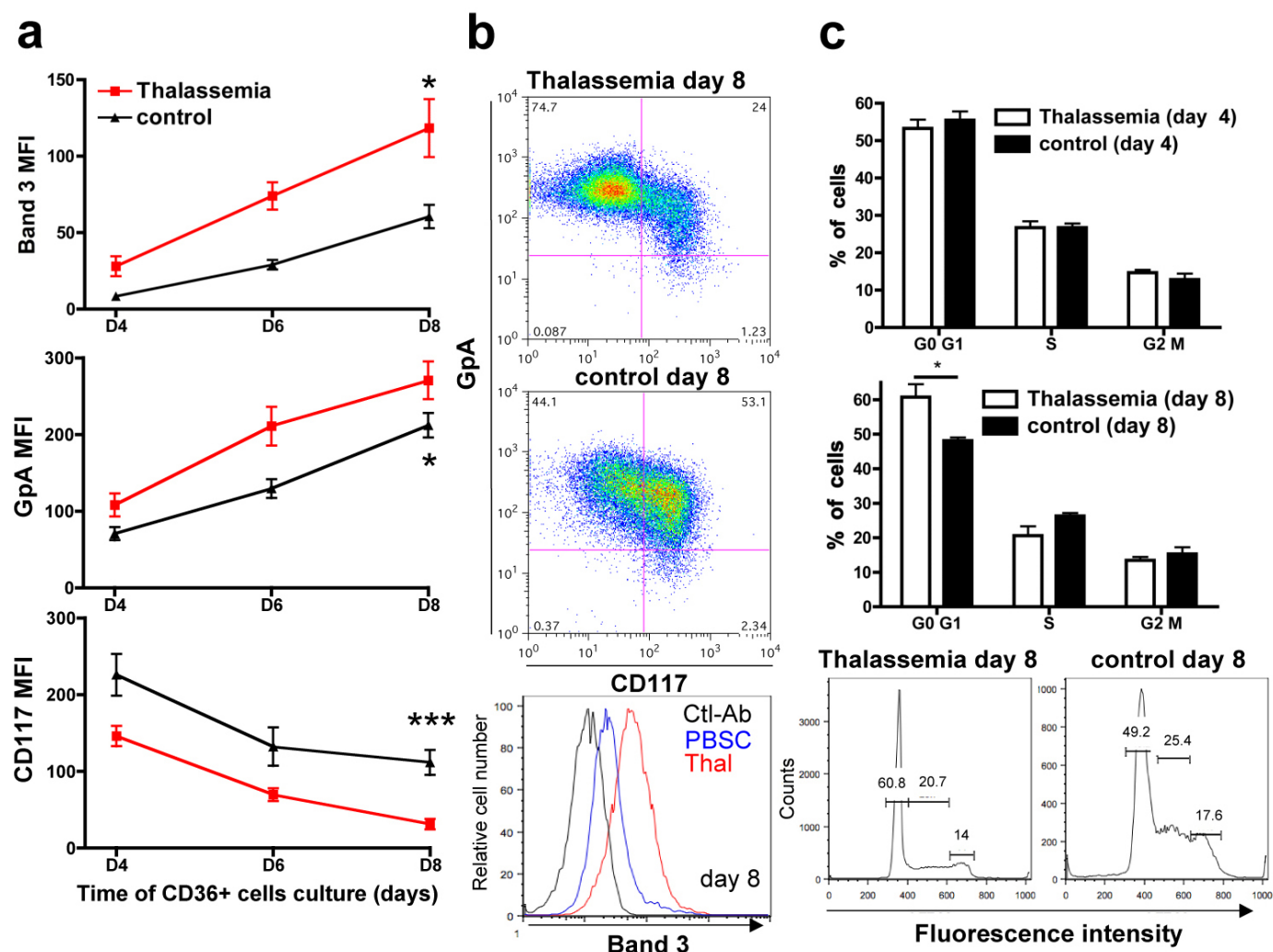
Retroviral production. Casp-3-uncleavable mutant GATA-1 (μ GATA-1) was cloned in a PINCO vector upstream of a cytomegalovirus-promoter–GFP cassette, which was a gift from R. De Maria. This construct was used to produce vector particles by

co-transfecting 293EBNA cells with the vector plasmid, an encapsidation plasmid (gag-pol) and an expression plasmid (pHCMV-G) encoding the vesicular stomatitis virus (VSVg) envelope, using jetPRIME transfection reagent.

Infection of erythroid cells. For the HSP70(S400A) and μ GATA-1 transduction, CD34⁺ cells isolated from β -TM peripheral blood mononuclear cells were cultured for 5 days, as described above. They were then infected by lentiviruses or retroviruses in the presence of 4 µg ml⁻¹ protamine sulphate. A second round of infection was performed 24 h later, upon addition of fresh medium with cytokines. After an additional 24 h, cells were extensively washed in PBS and stained with the anti-CD36-APC monoclonal antibody (BD Pharmingen). CD36⁺/GFP⁺ cells were then purified by flow cytometry cell sorting and differentiated in erythropoietin-containing medium. All data presented here were analysed at day 7 of the CD36⁺ cell culture. For the β -globin gene lentivirus transduction, CD34⁺ cells were isolated from β -TM peripheral blood mononuclear cells by magnetic sorting (MiltenyiBiotec). Sorted cells were pre-stimulated for 24 h in CellGro SCGM media supplemented with 100 ng ml⁻¹ stem cell factor, 100 ng ml⁻¹ thrombopoietin, 100 ng ml⁻¹ Flt3 ligand and 60 ng ml⁻¹ IL-3 at 37 °C and 5% CO₂. Next, pre-stimulated cells were transduced for 22 h with the LentiGlobin vector at an MOI (multiplicity of infection) of 50 in CellGro SCGM media supplemented with 100 ng ml⁻¹ stem cell factor, 100 ng ml⁻¹ thrombopoietin, 100 ng ml⁻¹ Flt3 ligand, 60 ng ml⁻¹ IL-3 and 4 µg ml⁻¹ protamine sulphate, or mock transduced in the same conditions. Two-phase liquid culture was then performed as described above.

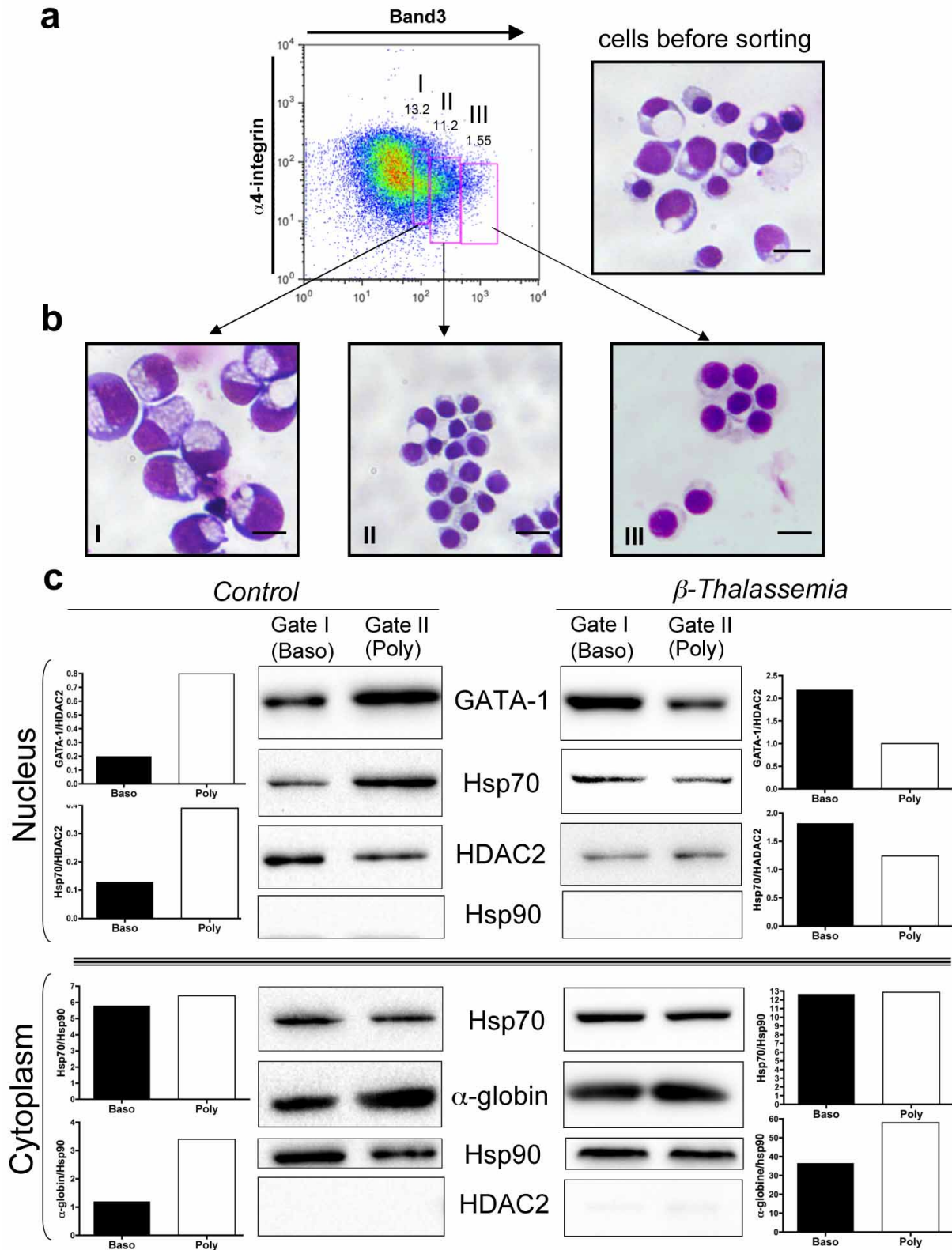
Statistical analyses. Statistical analyses were performed with GraphPadPrism (version 5.0; GraphPad Software). The data are expressed as the mean \pm standard deviation or median + IQR unless noted otherwise. Two-tailed Student's unpaired *t*-tests or Mann–Whitney *U*-tests were used as appropriate. *P* values <0.05 were considered significant, values smaller than this are indicated in figure legends.

23. De Franceschi, L. *et al.* K-CL co-transport plays an important role in normal and beta thalassemic erythropoiesis. *Haematologica* **92**, 1319–1326 (2007).
24. Zermati, Y. *et al.* Transforming growth factor inhibits erythropoiesis by blocking proliferation and accelerating differentiation of erythroid progenitors. *Exp. Hematol.* **28**, 885–894 (2000).
25. Gabet, A.-S. *et al.* Caspase-activated ROCK-1 allows erythroblast terminal maturation independently of cytokine-induced Rho signaling. *Cell Death Differ.* **18**, 678–689 (2011).
26. Bolte, S. & Cordelières, F. P. A guided tour into subcellular colocalization analysis in light microscopy. *J. Microsc.* **224**, 213–232 (2006).
27. Schallmeiner, E. *et al.* Sensitive protein detection via triple-binder proximity ligation assays. *Nature Methods* **4**, 135–137 (2007).
28. Kiger, L. *et al.* Dynamics of α -Hb chain binding to its chaperone AHSP depends on heme coordination and redox state. *Biochim. Biophys. Acta* **1840**, 277–287 (2014).
29. Abraham, E. C., Reese, A., Stallings, M. & Huisman, T. H. Separation of human hemoglobins by DEAE-cellulose chromatography using glycine-KCN-NaCl developers. *Hemoglobin* **1**, 27–44 (1976).
30. Bucci, E. & Fronticelli, C. A new method for the preparation of alpha and beta subunits of human hemoglobin. *J. Biol. Chem.* **240**, 551–552 (1965).
31. Anson, M. L. & Mirsky, A. E. Protein coagulation and its reversal: the preparation of insoluble globin, soluble globin and heme. *J. Gen. Physiol.* **13**, 469–476 (1930).



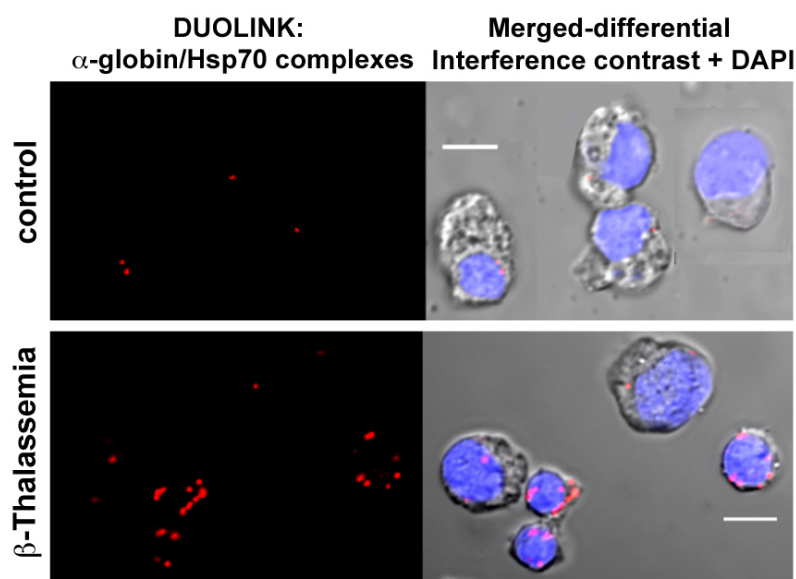
Extended Data Figure 1 | Kinetics of differentiation of β -TM and control erythroblasts and cell-cycle analysis. CD36⁺ cells derived from CD34⁺ adult β -TM or healthy (control) peripheral blood cells were cultured, as described in Methods. Differentiation staining was assessed at several points during CD36⁺ cell culture by flow cytometry. **a**, Differentiation curves determined using (top panel) band-3 MFI ($n = 4$ independent experiments), (middle panel) glycophorin-A (GpA) MFI and (bottom panel) CD117 MFI. For GpA and CD117, $n = 19$ independent experiments in β -TM group and $n = 11$ in control group. P values compare erythroid differentiation between β -TM and control cell cultures at day 8. **b**, Representative flow cytometry plots (GpA and

CD117 staining) showing difference in erythroid differentiation at day 8. GpA⁺/CD117⁺ cells represent mature erythroblasts. In the bottom panel, histograms of band-3 staining in β -TM and peripheral blood stem cells (PBSC, control erythroblasts) at day 8 are shown (Ctl-Ab, isotypic control antibody). **c**, Cell cycle was analysed at day 4 and day 8 of CD36⁺ cell cultures from β -TM and healthy donors ($n = 4$ independent experiments). Representative flow cytometry histograms at day 8 are shown in the bottom panel. Data in panels **a** and **c** are presented as mean \pm s.e.m. P values determined by two-tailed Student's unpaired t -tests. * $P < 0.05$, *** $P < 0.001$.



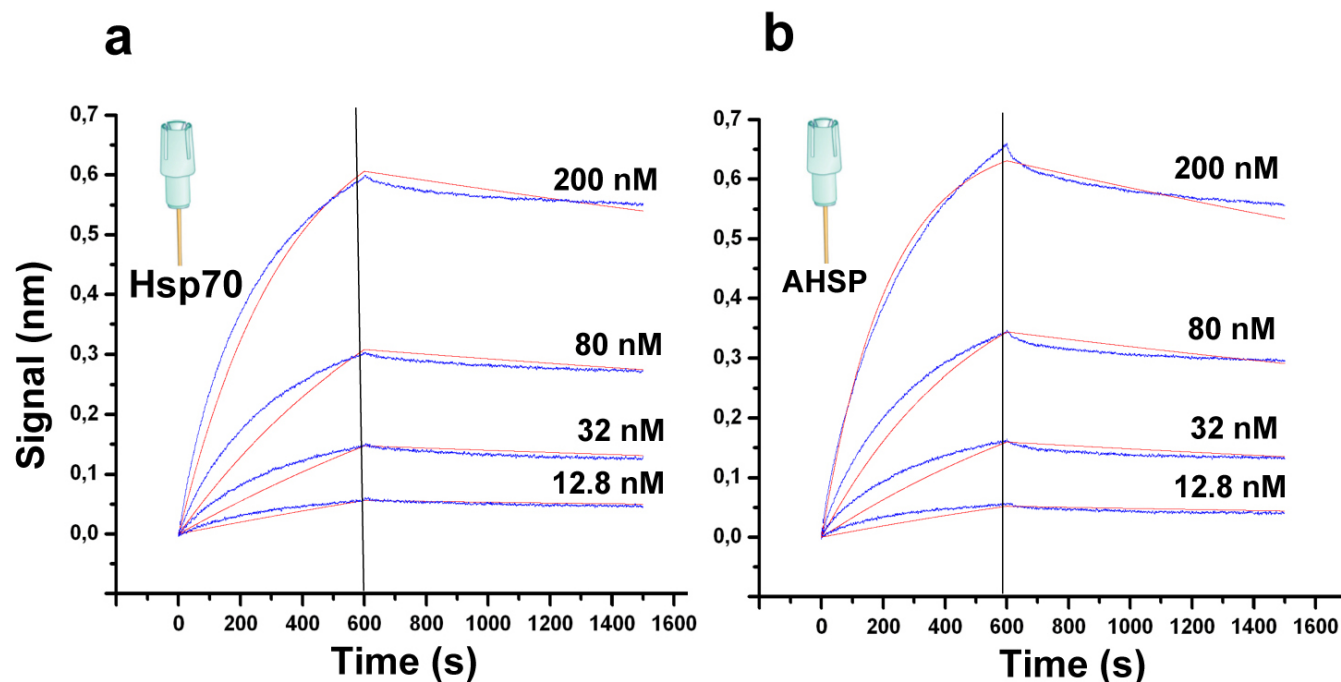
Extended Data Figure 2 | Specific developmental stage purification of erythroblasts and biochemical analysis of HSP70 and GATA-1 nuclear content. **a**, Strategy for flow cytometry (FACS) purification of erythroblasts at specific developmental stages at day 8 from CD36⁺ β -TM or control-derived cell cultures. Band-3 and $\alpha 4$ -integrin staining were used to isolate three stages of erythroid differentiation as described in Methods (representative β -TM-derived cell culture is shown). Numbers indicate the percentage of cells from each gate. Scale bar, 10 μ m. **b**, A representative morphological analysis ($\times 100$) of cells purified from each gate by MGG staining. Gate I, basophil erythroblasts

(Baso), gate II, polychromatophilic erythroblasts (Poly), gate III, acidophilic erythroblasts. Scale bar, 10 μ m. **c**, HSP70 and GATA-1 or α -globin western blot from nuclear (upper panel) or cytoplasmic (lower panel) fractions extracted from FACS-purified gate I and II β -TM or control-derived cell cultures (day 8). HDAC2 and Hsp90 were used as controls for the nuclear and cytoplasmic fractions, respectively. Graphs show optical density values of specific protein normalized to that of HDAC2 or Hsp90. Data are representative of three independent experiments.



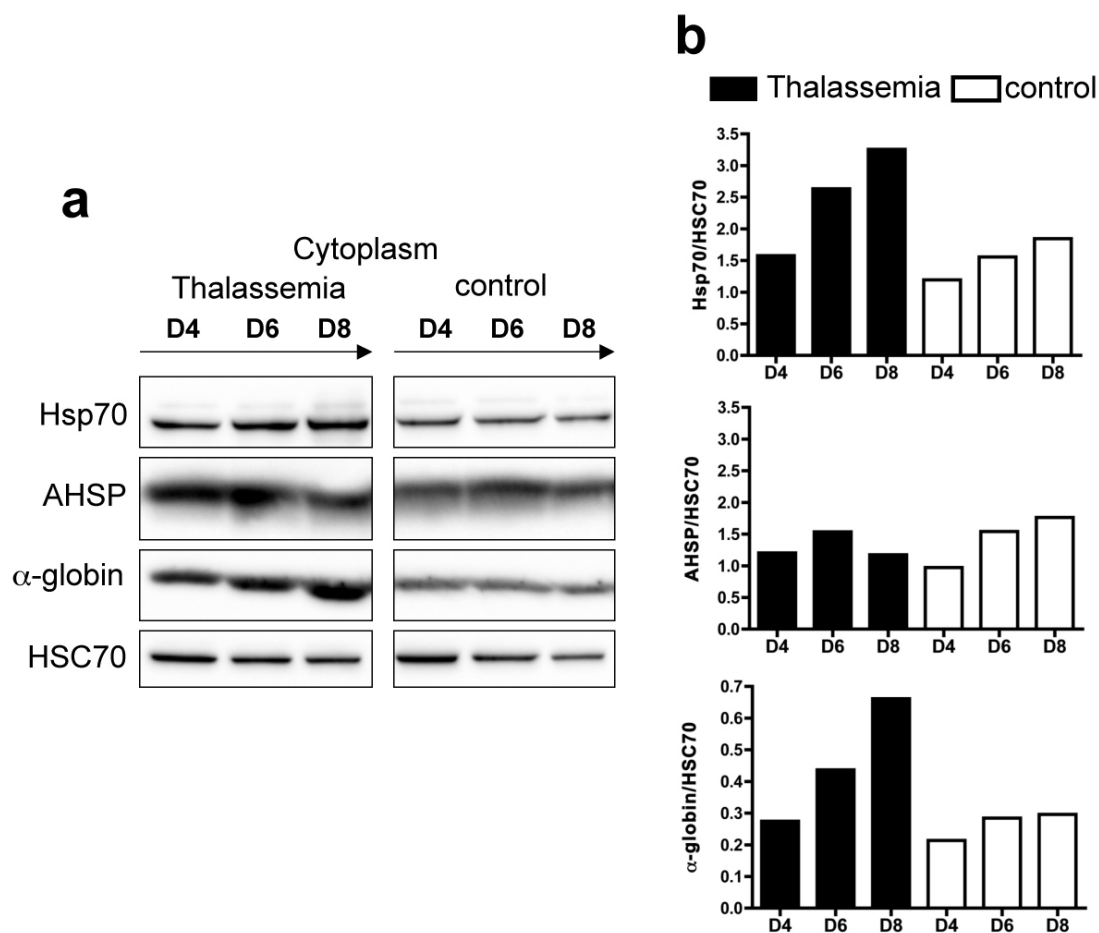
Extended Data Figure 3 | HSP70- α -globin complexes correlate with the degree of cell differentiation. Protein interactions were analysed at day 8 from CD36⁺ β -TM or control-derived cell cultures by Duolink *in situ* proximity ligation assay, using HSP70 and α -globin antibodies. The red spots indicate

close proximity between cellular-bound antibodies. More spots per cell are exhibited in more differentiated β -TM erythroblasts (smaller cells). A representative experiment is shown ($n = 3$). Nuclei are stained with DAPI (blue). Scale bar, 5 μ m.



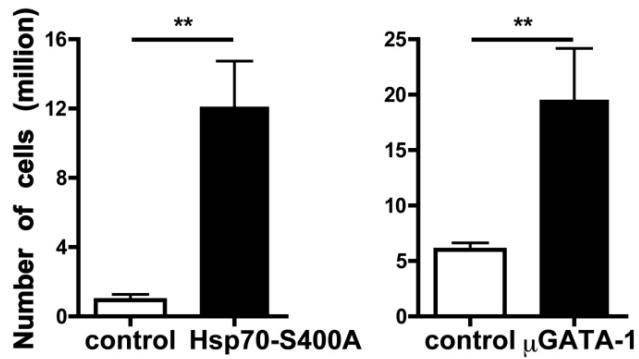
Extended Data Figure 4 | Comparative study of *apo-α*-globin association to HSP70 or to AHSP by biolayer interferometry. Sensorgrams are shown in blue and the corresponding fits in red. The data produced a K_d of 4.65 nM and 8.35 nM for HSP70-*apo-α*-globin (a) and AHSP-*apo-α*-globin

(b) interactions, respectively. χ^2 was below 1.3 and R^2 was above 0.99 for all included sensorgrams. Data are representative of two independent experiments.

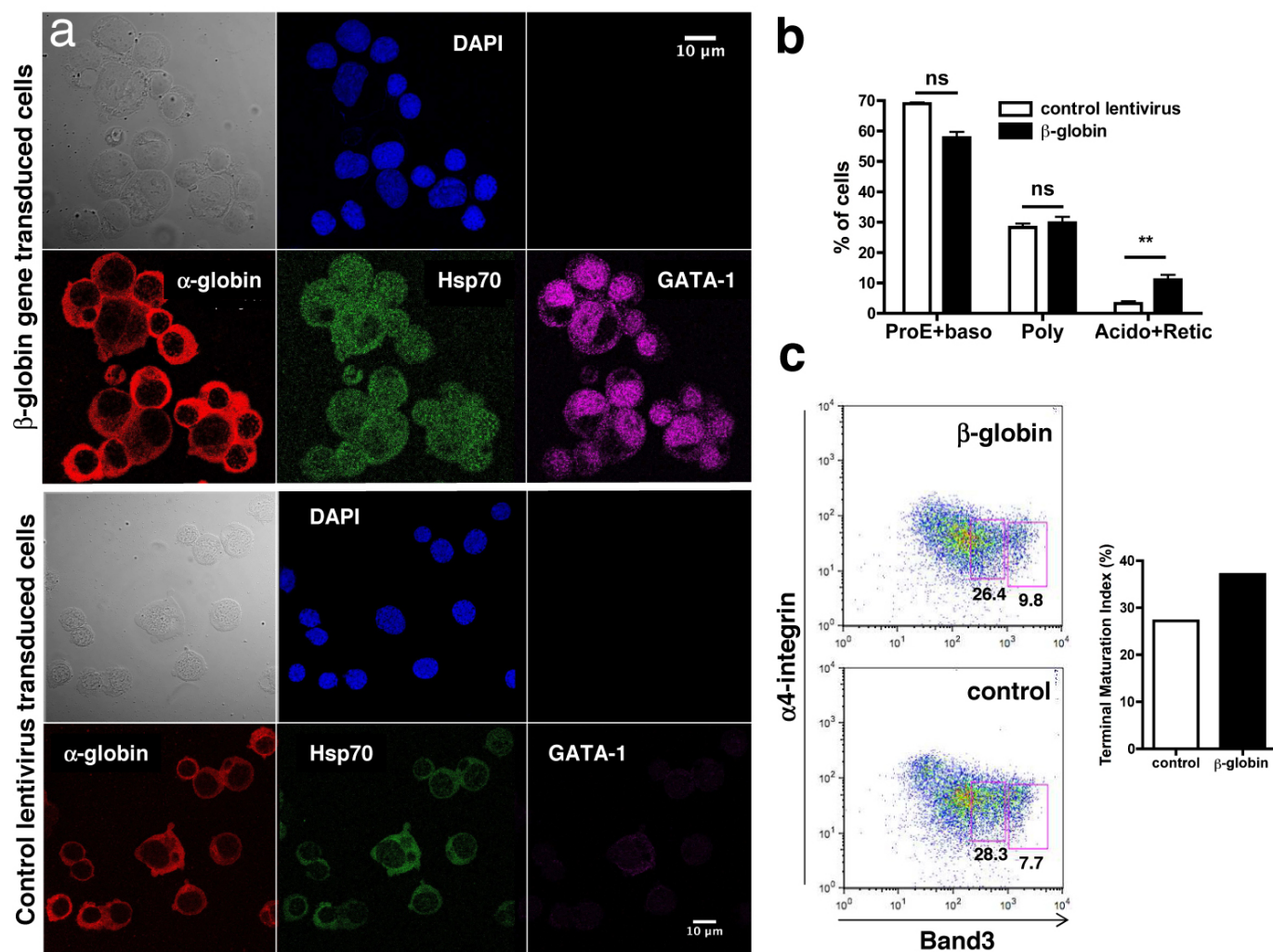


Extended Data Figure 5 | Kinetics of cytoplasmic HSP70 and AHSP expression during erythropoiesis. CD36⁺ cells derived from CD34⁺ adult β -TM or healthy (control) peripheral blood cells were cultured, as described in Methods section. **a**, **b**, HSP70, AHSP and α -globin expression was assessed in cytoplasm at several culture times (days 4, 6 and 8 of the CD36⁺ cell culture)

by western blot analysis. **a**, A representative western blot is shown ($n = 3$ independent experiments). β -TM- and control-cell lysates have been loaded in the same gel. **b**, Graphs show the optical density values of specific proteins normalized to cytoplasmic protein HSC70.



Extended Data Figure 6 | Absolute number of very mature erythroblasts increases with nuclear HSP70 and GATA-1 restoration. Comparison of the absolute number of very mature erythroblasts (acidophilic cells + reticulocytes) at day 7 of CD36⁺ cell culture after transduction of β -TM CD34⁺ cells with a nuclear-targeted HSP70 lentiviral mutant (HSP70(S400A)) or a retroviral mutant of GATA-1 uncleavable by activated caspase-3 (μ GATA-1) and their empty vector as control. Data are presented as mean \pm s.e.m. for three independent experiments for each transduction. *P* values were determined using Student's unpaired *t*-test. ***P* < 0.01.



Extended Data Figure 7 | The lentiviral transduction of β -TM CD34⁺ cells with the β -globin gene rescues HSP70 nuclear localization and GATA-1 protection. β -TM CD34⁺ cells derived from peripheral blood cells were infected with a lentivirus encoding the β -globin gene or a control lentivirus ($n = 2$, 10–15% transduction efficiency). They were then cultured with a two-phase amplification liquid culture system, as described in Methods. Results at day 8 of CD36⁺ cell culture are shown. **a**, Confocal microscopy analysis of HSP70 nuclear localization and GATA-1 expression for each transduction.

Scale bar, 10 μ m. **b**, **c**, Erythroid differentiation. The graph in **b** represents the proportion of cells at each developmental stage after MGG staining for each transduction (mean \pm s.e.m.). P values were determined by two-tailed Student's unpaired t -tests. ** $P < 0.01$; ns, not significant. Representative flow cytometry plots of band-3 and α_4 -integrin staining are shown in **c**, left panel. The terminal maturation index of this experiment as calculated by flow cytometry is shown in **c**, right panel.

PRC2 loss amplifies Ras-driven transcription and confers sensitivity to BRD4-based therapies

Thomas De Raedt^{1,2,3}, Eline Beert^{4*†}, Eric Pasmant^{5,6*}, Armelle Luscan^{5,6}, Hilde Brems⁴, Nicolas Ortonne^{5,6}, Kristian Helin^{7,8,9}, Jason L. Hornick¹⁰, Victor Mautner¹¹, Hildegard Kehrer-Sawatzki¹², Wade Clapp¹³, James Bradner^{2,14}, Michel Vidaud^{5,6}, Meena Upadhyaya¹⁵, Eric Legius^{4,16} & Karen Cichowski^{1,2,3}

The polycomb repressive complex 2 (PRC2) exerts oncogenic effects in many tumour types¹. However, loss-of-function mutations in PRC2 components occur in a subset of haematopoietic malignancies, suggesting that this complex plays a dichotomous and poorly understood role in cancer^{2,3}. Here we provide genomic, cellular, and mouse modelling data demonstrating that the polycomb group gene *SUZ12* functions as tumour suppressor in PNS tumours, high-grade gliomas and melanomas by cooperating with mutations in *NF1*. *NF1* encodes a Ras GTPase-activating protein (RasGAP) and its loss drives cancer by activating Ras⁴. We show that *SUZ12* loss potentiates the effects of *NF1* mutations by amplifying Ras-driven transcription through effects on chromatin. Importantly, however, *SUZ12* inactivation also triggers an epigenetic switch that sensitizes these cancers to bromodomain inhibitors. Collectively, these studies not only reveal an unexpected connection between the PRC2 complex, *NF1* and Ras, but also identify a promising epigenetic-based therapeutic strategy that may be exploited for a variety of cancers.

The Ras pathway is commonly deregulated in cancer⁵. Oncogenic *RAS* mutations occur in approximately 30% of tumours; however, mutations in upstream regulators and downstream effectors are also prevalent⁵. Irrespective of aetiology, the oncogenic Ras signal is frequently potentiated as cancers progress, in some instances through amplification of mutant *RAS* genes or suppression of negative feedback pathways^{6–9}. It is likely that many unknown mechanisms contribute to this emerging aspect of tumour evolution.

Mutations in RasGAP genes promote Ras activation in many cancers^{4,10,11}. Defects in the *NF1* RasGAP underlie a familial cancer syndrome (NF1), sporadic glioblastomas (GBMs), neuroblastomas and lung cancer⁴. Interestingly, some severely affected patients with NF1 (NF1 microdeletion patients) carry germline deletions of *NF1* and 13 surrounding genes^{12,13}. These individuals develop thousands of benign tumours and can exhibit a fourfold increased risk of malignancies^{14,15}, suggesting that a cooperating tumour suppressor lies within this region. If true, this gene might affect the development of tumours in patients with NF1 and in sporadic cancers, which frequently possess large 17q deletions¹⁶. Therefore, we analysed 51 malignant peripheral nerve sheath tumours (MPNSTs), the most common NF1-associated malignancy, from patients with and without an NF1 microdeletion. Array comparative genomic hybridization revealed heterozygous and homozygous deletions of *SUZ12*, a gene within the microdeletion, which encodes a component of the PRC2 (Fig. 1a, Extended Data Fig. 1a and Extended Data Table 1). Deletions in *EED*, another PRC2 component, were also prevalent in both populations.

Moreover, inactivating mutations in *SUZ12* and *EED* were frequently present (Fig. 1a, Extended Data Fig. 1b, c and Supplementary Information). Together, 79% of microdeletion and 34% of non-microdeletion tumours exhibited homozygous loss of *SUZ12* or *EED*; however, 53% of non-microdeletion MPNSTs harboured one or more PRC2 mutations/deletions (Fig. 1b). These observations suggest that PRC2 inactivation may play a critical role in MPNST development.

SUZ12 and H3K27me3 were undetectable in primary MPNSTs harbouring *SUZ12* mutations and in four *NF1*-deficient human MPNST cell lines (Extended Data Fig. 1d, e and Fig. 1c). *SUZ12* reconstitution restored H3K27me3, reduced proliferation and triggered cell death in low serum (Fig. 1c, d and Extended Data Fig. 1f). The *SUZ12* lentivirus did not affect *SUZ12* wild-type cells (Extended Data Fig. 1g). Conversely, *SUZ12* ablation enhanced colony growth of *NF1*-deficient (U251 (ref. 17)) but not *NF1* wild-type GBM cells (Fig. 1e and Extended Data Fig. 1h). Partly effective short hairpin RNAs (shRNAs) exerted a disproportionate suppressive effect on H3K27me3 and promoted colony growth of U251 cells, albeit less efficiently (Fig. 1e). These results suggest that reduced PRC2 dosage might also contribute to tumour development, which could be relevant in the setting of heterozygous *SUZ12/EED* mutations (Fig. 1b). Finally, *SUZ12* ablation enhanced colony growth of *NF1*-deficient but not *NF1* wild-type melanoma cells (Fig. 1f and Extended Data Fig. 1i).

To assess the cooperativity of these genes further, *Nf1*^{+/-}; *Suz12*^{+/-} 'cis' mice, carrying mutations on the same chromosome, were generated to recapitulate the genetics of microdeletion patients and spontaneous tumours. Tumours in *Nf1* mouse models are driven by spontaneous loss of wild-type chromosome 11 (ref. 18), resulting in *Nf1* and *Suz12* loss (Extended Data Fig. 2a). *Nf1*^{+/-}; *Suz12*^{+/-} mice were highly tumour prone and exhibited decreased survival (Fig. 2a). Notably, mice developed lesions associated with both germline and spontaneous *NF1* mutations in humans, including neurofibromas, MPNSTs, schwannomas, melanocytic nevi and intestinal adenomas (Fig. 2b and Extended Data Fig. 2b)⁴. Genomic human tumour data demonstrate that at least one allele of *NF1* and *SUZ12* are co-deleted/mutated in 24% of melanomas (Fig. 2c). Moreover, *EED* deletions/mutations are present in 44% of *NF1*-mutant tumours, most of which also lack one *SUZ12* allele.

To investigate *SUZ12* in MPNST development further, mice harbouring *cis* mutations in *Nf1*, *p53* and *Suz12* were also generated, as combined *NF1/p53* mutations drive MPNST formation^{18–20} (Supplementary Information). These *Nf1/p53/Suz12* mice exhibited decreased survival and developed MPNSTs on average 2.3 months earlier than *Nf1/p53* mutant

¹Genetics Division, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. ²Harvard Medical School, Boston, Massachusetts 02115, USA. ³Ludwig Center at Dana-Farber/Harvard Cancer Center, Boston, Massachusetts 02115, USA. ⁴Department of Human Genetics, Catholic University Leuven, 3000 Leuven, Belgium. ⁵INSERM UMR_S 745 et EA7331, Université Paris Descartes, Sorbonne Paris Cité, Faculté des Sciences Pharmaceutiques et Biologiques, 75006 Paris, France. ⁶Service de Biochimie et Génétique Moléculaire, Hôpital Cochin, Assistance Publique-Hôpitaux de Paris, 75014 Paris, France. ⁷Biotech Research and Innovation Centre (BRIC), University of Copenhagen, 2200 Copenhagen, Denmark. ⁸Center for Epigenetics, University of Copenhagen, 2200 Copenhagen, Denmark. ⁹The Danish Stem Cell Center (Danstem), University of Copenhagen, 2200 Copenhagen, Denmark. ¹⁰Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. ¹¹Department of Maxillofacial Surgery, University Medical Centre, Hamburg-Eppendorf, 20246 Hamburg, Germany. ¹²Institute of Human Genetics, University of Ulm, 89081 Ulm, Germany. ¹³Herman Wells Center for Pediatric Research, Department of Pediatrics, Indiana University School of Medicine, 46202 Indianapolis, Indiana, USA. ¹⁴Department of Medical Oncology, Dana-Farber Cancer Institute, Massachusetts 02115, USA. ¹⁵Institute of Medical Genetics, Cardiff University, Heath Park, Cardiff CF14 4XN, UK. ¹⁶Center for Human Genetics, University Hospital Leuven, 3000 Leuven Belgium. [†]Present address: Laboratory of Aquatic Biology, Interdisciplinary Research Facility Life Sciences, Katholieke Universiteit, Leuven Afdeling Kortrijk, 8500 Kortrijk, Belgium.

*These authors contributed equally to this work.

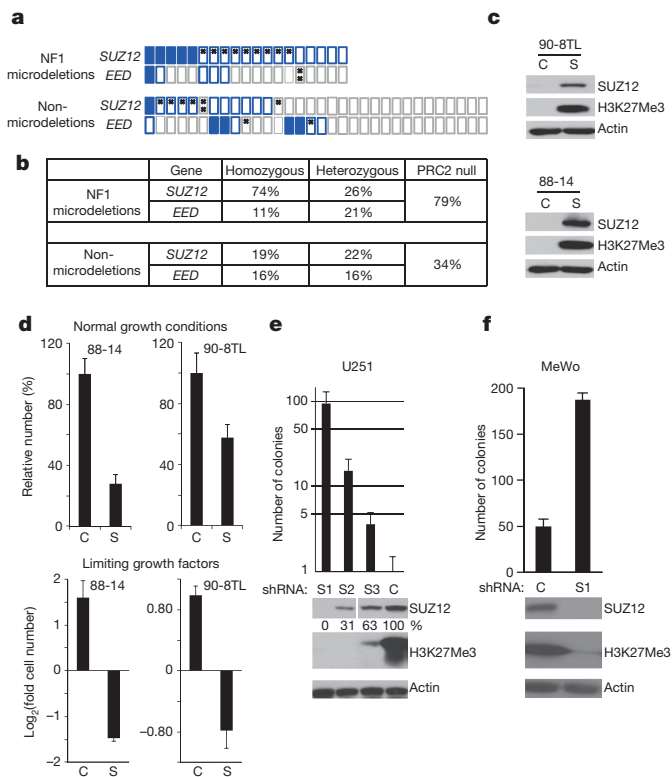


Figure 1 | PRC2 components are deleted or mutated in human MPNSTs and exert tumour-suppressive activity. **a**, Schematic visualization of distinct genomic alterations in *SUZ12* and *EED* in *NF1*-deficient human MPNSTs from patients with and without an *NF1* microdeletion (asterisk, mutation; open blue rectangle, hemizygous deletion; filled blue rectangle, homozygous deletion; grey rectangle, no copy number loss). **b**, Table indicating the percentage of human MPNSTs with homozygous or heterozygous loss of *SUZ12*, *EED* and *PRC2*. **c**, Immunoblots of lysates from human MPNST cells reconstituted with *SUZ12* (C, LacZ control; S, *SUZ12*). Restoration of H3K27me3 shows that the functionality of the PRC2 complex is restored upon *SUZ12* reconstitution. **d**, Human MPNST cell lines with reconstituted *SUZ12* exhibit reduced proliferation under normal growth conditions (10% serum, 3 days). *SUZ12* reconstitution (S) causes cell death when cells are cultured under limiting growth factors (2% serum, 6 days). Proliferation curves are shown in Extended Data Fig. 1f. **e**, Top, graph depicting colony growth in soft agar using the *NF1*-deficient human GBM line (U251), expressing a control shRNA construct (C, short hairpin green fluorescent protein (shGFP)) or three distinct *SUZ12* shRNA constructs against *SUZ12* (S1–S3). Bottom, immunoblots depicting knockdown of *SUZ12* to varying levels (S1–S3) compared with control (shGFP). The percentage of *SUZ12* remaining is quantified. **f**, Effects of sh*SUZ12* (S1) on *SUZ12* expression and colony formation in MEWO melanoma cells, which are *NF1* null. Error bars, s.d. ($n = 3$, biological replicates).

mice developed high-grade gliomas and did so on average in approximately 3 months, a latency shorter than most models (Fig. 2e, f and Extended Data Fig. 2d). *NF1* and *SUZ12* were co-deleted/mutated in 14% of human GBMs (Fig. 2g). *EED* deletions were also detected in 26% of *NF1*-mutant tumours, which all lacked a copy of *SUZ12*, highlighting the co-occurrence of defects in *NF1* and PRC2.

SUZ12 ablation causes a loss of H3K27me3 and in some settings promotes a consequential increase in H3K27Ac²¹: a transcriptional activating signal that recruits bromodomain proteins and associated transcription factors²². As such, we reasoned that *SUZ12*-mutant MPNSTs might be sensitive to bromodomain inhibitors such as JQ1²³. Importantly, *SUZ12* reconstitution increased H3K27me3 and decreased H3K27Ac in MPNST

mice (Fig. 2d, e). All MPNSTs arose in less than 3.5 months, whereas MPNSTs in *Nf1/p53* mice required up to 9 months. Nevertheless, fewer *Nf1/p53/Suz12* animals developed MPNSTs, most probably because of the increase in other tumour types (Fig. 2f). Strikingly, 54% of *Nf1/p53/Suz12*

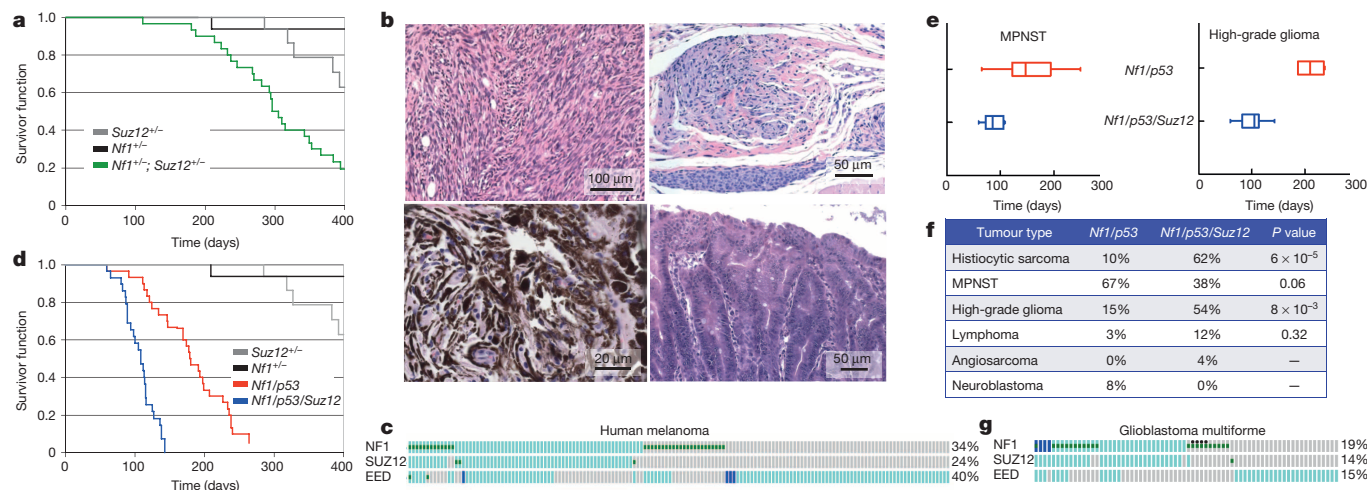


Figure 2 | *Suz12* and *Nf1* mutations cooperate to promote widespread tumour development in mice. **a**, Kaplan–Meier curve comparing survival of *Suz12*^{+/−} ($n = 15$), *Nf1*^{+/−} ($n = 15$) and *Nf1*^{+/−}; *Suz12*^{+/−} ($n = 30$) mice in cis (NS^{+/−}). *Nf1*^{+/−}; *Suz12*^{+/−} mice have a significantly reduced survival (*Nf1*^{+/−}; *Suz12*^{+/−} versus *Suz12*^{+/−}, $P = 1.99 \times 10^{-4}$; *Nf1*^{+/−}; *Suz12*^{+/−} versus *Nf1*^{+/−}, $P = 2.87 \times 10^{-8}$; Mantel–Cox test). **b**, Haematoxylin and eosin staining of an MPNST (top left), neurofibroma (top right), melanocytic nevus (bottom left) and intestinal adenoma (bottom right) in *Nf1*^{+/−}; *Suz12*^{+/−} mice. **c**, Oncoprint output of the melanoma TCGA data (cBioPortal) (green square, mutation; light blue rectangle, hemizygous deletion; dark blue rectangle, homozygous deletion; grey rectangle, unaltered). **d**, Kaplan–Meier curve comparing survival of *Suz12*^{+/−}, *Nf1*^{+/−}, *Nf1*^{+/−}; *p53*^{+/−} mice ($n = 30$) in cis (NP^{+/−}) and *Nf1*^{+/−}; *p53*^{+/−}; *Suz12*^{+/−} mice in cis (NPS^{+/−}). *Nf1*^{+/−}; *p53*^{+/−}; *Suz12*^{+/−} mice have a significantly lower survival ($P = 1.2 \times 10^{-9}$, Mantel–Cox

test). **e**, Left, box plot comparing the age at onset for MPNSTs in *Nf1/p53* and *Nf1/p53/Suz12* mice. MPNSTs from *Nf1/p53/Suz12* mice have a significantly shorter latency (on average 2.3 months; $P = 2.8 \times 10^{-4}$, Mann–Whitney *U*-test). Right, box plot comparing the age at onset for high-grade glioma in *Nf1/p53* and *Nf1/p53/Suz12* mice. High-grade gliomas in *Nf1/p53/Suz12* mice develop on average around 3 months and have a significantly shorter latency ($P = 0.0003$, Mann–Whitney *U*-test). **f**, Table comparing the percentages of *Nf1/p53* and *Nf1/p53/Suz12* mice that develop specific tumours. *P* values indicate significance by Fisher's exact test. **g**, Oncoprint output of the glioblastoma TCGA data (cBioPortal) (green square, mutation; light blue rectangle, hemizygous deletion; dark blue rectangle, homozygous deletion; black dot, second hit mutation). Co-occurrence of defects in *NF1* and the PRC2 complex (*NF1*–*SUZ12*, odds ratio >10 , $P < 10^{-4}$; *NF1*–*EED*, odds ratio 2–10, $P = 2.2 \times 10^{-4}$).

cells, whereas *SUZ12* ablation decreased H3K27me3 and increased H3K27Ac (Fig. 3a). Moreover, *SUZ12*-deficient MPNSTs were more sensitive to JQ1 than PRC2 wild-type cells (Fig. 3b).

Transcriptional profiling analysis was conducted in MPNSTs, in the presence/absence of *SUZ12* or JQ1. Gene set enrichment analysis (GSEA) and single-sample (ss)GSEA demonstrated that the gene set suppressed by *SUZ12* was also suppressed by JQ1 (Supplementary Table 1 and Fig. 3c). Using the Molecular Signatures Database we found that PRC2-regulated genes were suppressed both in *SUZ12*-reconstituted and JQ1-treated cells, suggesting that JQ1 counteracts *SUZ12* loss, in part, by repressing PRC2 targets (Extended data Fig. 3a, b). Unlike other tumours, neither JQ1 nor *SUZ12* suppressed *MYC* genes (Extended Data Fig. 3c). However, *MYC* genes are not upregulated and do not drive MPNST development. Surprisingly, *SUZ12* and JQ1 also suppressed genes associated with oncogenic Ras transcriptional signatures (Fig. 3d and Extended Data Fig. 3a). Conversely, Ras signatures were enhanced when *SUZ12* was ablated in *NF1*-mutant GBM cells (Fig. 3d). Phospho-ERK was unaffected by *SUZ12* loss, reconstitution, or JQ1 treatment (Extended Data Fig. 3d), suggesting that *SUZ12* loss amplifies Ras transcriptional output via direct effects on chromatin, and that JQ1 may counteract effects by suppressing Ras signature genes.

There are no effective therapies for MPNSTs²⁴. While *SUZ12*-mutant MPNSTs were sensitive to JQ1, in contrast to c-Myc driven tumours, the effects were cytostatic^{25,26}. Given the importance of Ras in MPNSTs, we investigated whether JQ1 might cooperate with Ras pathway inhibitors.

JQ1 was evaluated in combination with an mTOR inhibitor (rapamycin), or a MEK inhibitor (PD-901) as both effectors function in *NF1*-deficient tumours^{27–30}. Only PD-901/JQ1 treatment killed MPNSTs (Fig. 4a). Genetic ablation of BRD4 similarly cooperated with PD-901 (Extended Data Fig. 4a, b). *SUZ12* suppression conferred sensitivity to PD-901/JQ1 in other *NF1*-deficient tumour cells, including GBM and colon cancer, whereas *SUZ12* suppression did not affect *NF1* wild-type cells (Extended Data Fig. 4c and Supplementary Information).

To elucidate the cooperativity between BRD4 and MEK inhibitors, transcriptional profiling and chromatin immunoprecipitation followed by sequencing (ChIP-seq) experiments were performed. We compiled a list of publicly available upregulated Ras signature genes significantly repressed by JQ1, PD-901 or JQ1/PD901 in MPNSTs (Supplementary Table 2 and Supplementary Information). PD-901 suppressed a high percentage (58%) of these Ras signature genes (Fig. 4b, blue boxes) and JQ1 suppressed 29% (yellow boxes). However, JQ1/PD-901 cooperated to suppress this Ras signature in several ways. Not only did JQ1 suppress genes unaffected by PD-901, but the JQ1/PD-901 combination enhanced or uniquely suppressed 26% of these genes (green boxes). ssGSEA of our Ras signature gene set confirmed the cooperative effects of PD901 and JQ1 (Fig. 4c). H3K27Ac and BRD4 were significantly enriched at promoters of JQ1/PD-901 suppressed targets (Fig. 4b, red boxes) and BRD4 was displaced at 53% of sites (red boxes with dots), as shown in Fig. 4d. Transcription factors including ETS1, a known master regulator of Ras-driven transcription, were suppressed by JQ1/PD-901.

Finally, PD-901 and JQ1 were evaluated in *Nf1/p53/Suz12* mice (Fig. 4e). PD-901 did not affect MPNSTs, whereas JQ1 exerted modest, cytostatic effects. However, combined JQ1/PD-901 triggered significant tumour regression. On average, tumours shrank 67%; however, half of them shrank by at least 75%. On the basis of their mechanism of development, tumours from *Nf1/p53* mice are *SUZ12* heterozygous, recapitulating an additional subset of human tumours (Fig. 1b). Consistent with shRNA studies (Fig. 1e), tissue from *Suz12*^{+/-} mice revealed a disproportionate loss of H3K27me3 (Fig. 4f), potentially conferring sensitivity to PD-901/JQ1. Indeed, *Nf1*^{-/-}; *p53*^{-/-}; *Suz12*^{+/-} tumours shrank up to 68% (Fig. 4g), albeit overall less than tumours in *Nf1/p53/Suz12* mice. Notably, H3K27me3 antibodies faithfully detected PRC2 defects in primary human tumours (Fig. 4h, Extended Data Fig. 4d and

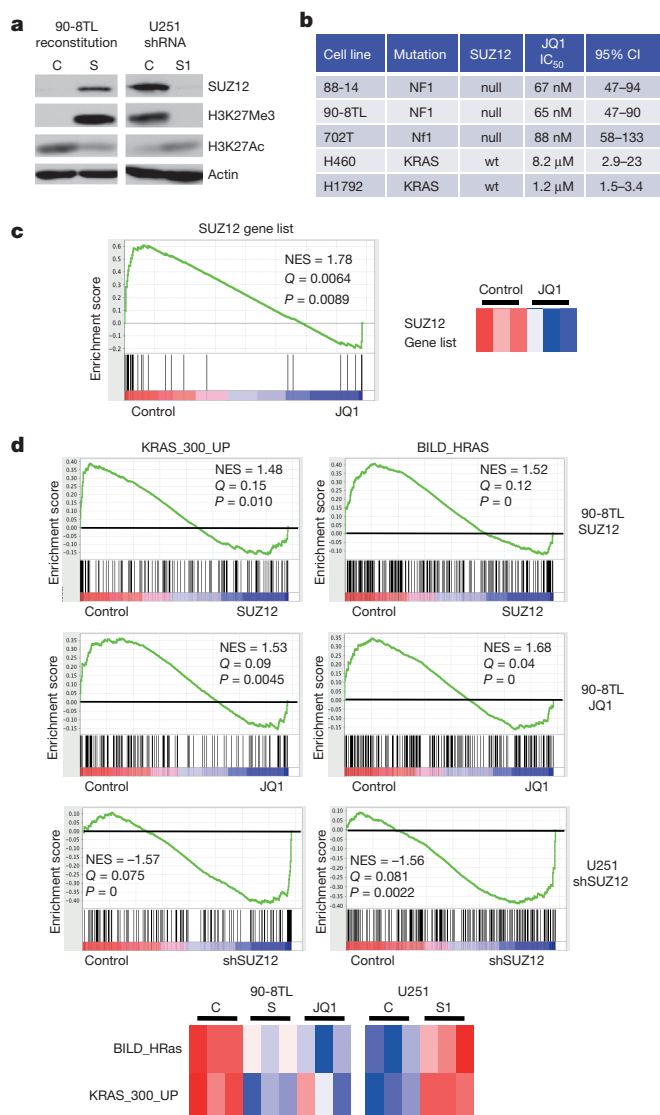
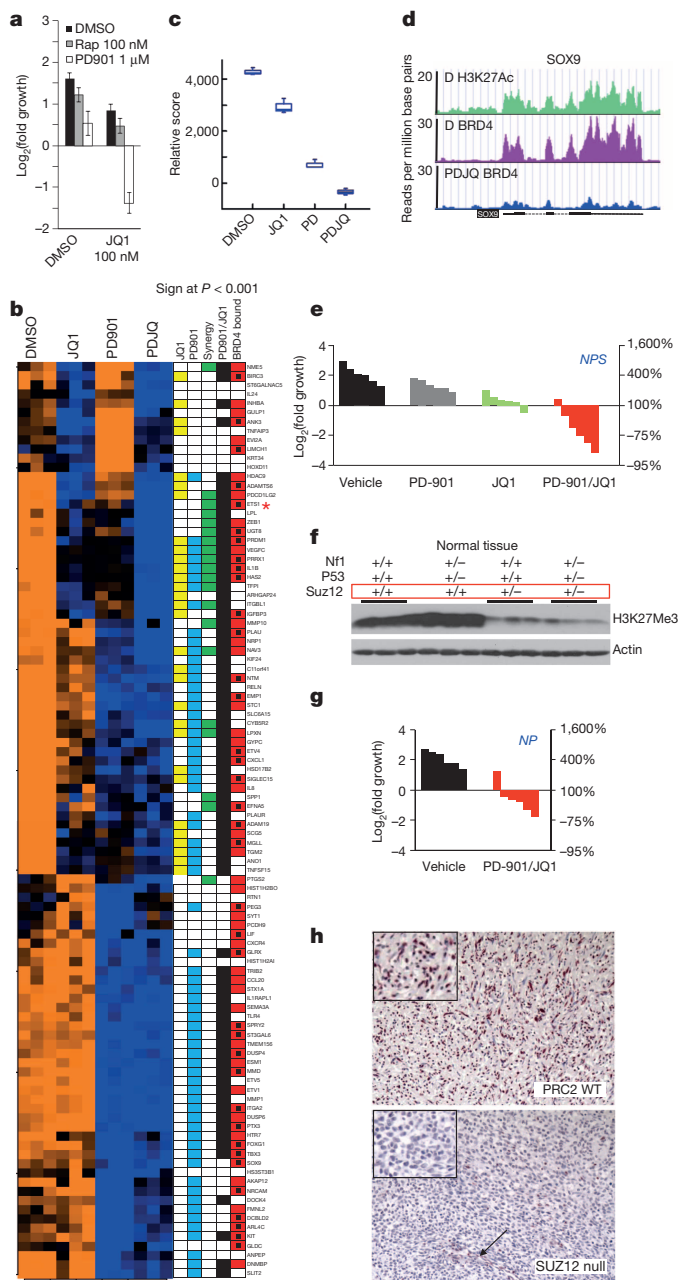


Figure 3 | *SUZ12* and JQ1 regulate PRC2 targets and the Ras transcriptional signature. **a**, Left, immunoblots showing the epigenetic switch from H3K27Ac to H3K27Me3 upon *SUZ12* reconstitution in human 90-8TL MPNST cells (C, LacZ; S, *SUZ12* reconstitution; H3K27Me3 expression included here with immunoblots is also shown in Fig. 1c). Right, epigenetic switch from H3K27Me3 to H3K27Ac upon *SUZ12* knockdown in *NF1*-deficient GBMs (U251) (C, scramble shRNA; S1, *SUZ12* shRNA). **b**, Half-maximum inhibitory concentration (IC₅₀) values of JQ1 in cell lines null and wild type (WT) for *SUZ12* and *NF1* as indicated. 702T cells are *Nf1/p53/suz12* null MPNSTs derived from our mouse tumours. KRAS mutant/*SUZ12* WT H460 and H1792 lung cancer lines were used as a comparison. 95% CI, 95% confidence interval. **c**, Top, signature enrichment plot comparing the *SUZ12* signature (genes downregulated in response to *SUZ12* reconstitution; Supplementary Table 1) in MPNST cells treated with vehicle versus JQ1 (100 nM, 24 h, triplicate samples). Plot indicates that *SUZ12*-regulated genes are more highly expressed in control than JQ1-treated samples (red, upregulation; blue, downregulation). NES, normalized enrichment score. Bottom, heat map of the enrichment scores of *SUZ12* downregulated genes in vehicle (control) and JQ1 (three replicates) (red, upregulation; blue, downregulation). **d**, Top, Ras signature enrichment plots (using KRAS_300_UP.V1_UP and BILD_HRAS signature data sets obtained from the Molecular Signatures Database) in MPNSTs (90-8TL) reconstituted with *SUZ12* versus control vector or treatment with JQ1 versus vehicle control (100 nM, 24 h, triplicate samples) and in the human GBM line U251 after knockdown of *SUZ12* (shSUZ12 S1). Plots indicate a significant downregulation of Ras signatures after *SUZ12* reconstitution and JQ1 treatment, and a significant upregulation of Ras signatures in *SUZ12*-depleted cells. Bottom, heat map of the enrichment scores of Ras signatures in control (C, LacZ), *SUZ12*-reconstituted (S) and JQ1-treated MPNST cells (three replicates), and in control (C, shScramble) and *SUZ12*-ablated (S1, shSUZ12) U251 cells (three replicates) (red, upregulation; blue, downregulation).



Extended Data Table 1). Together, these studies reveal a promising therapeutic strategy for MPNSTs, and suggest that PRC2 genetic alterations or H3K27me3/H3K27Ac levels could be used as biomarkers for selection of patients.

While the PRC2 complex is oncogenic in many cancers, these studies reveal an important tumour-suppressive function for two PRC2 components in *NF1*-deficient solid tumours. Moreover, this complex appears to function as a rheostat for Ras-dependent transcription. It is becoming apparent that the oncogenic Ras signal is amplified as cancers progress. Because *NF1*-mutant tumours are driven by a relatively weak oncogenic signal and lack an amplifiable mutant *RAS* allele, *SUZ12* loss may be a particularly important mechanism of signal potentiation. PRC2 inactivation may also enhance Ras-driven transcription in *RAS*-mutant cancers. However, further studies will be required to assess this possibility.

Additionally, these studies demonstrate that *SUZ12* loss promotes an epigenetic switch from H3K37Me3 to H3K27Ac, conferring sensitivity to BRD4 inhibitor-based combination therapies. Our data suggest that

Figure 4 | JQ1 and MEK inhibitors cooperate to promote cell death, suppress Ras transcriptional output and promote tumour regression.

a, Log₂ of fold change in MPNST cell number (90-8TL). Only cells treated with PD-901 and JQ1 die. Error bars, s.d. ($n = 3$, biological replicate). **b**, Heat map of microarray expression data evaluating the effects of the indicated drugs in *SUZ12/NF1* null cell line 90-8TL (Ras signature gene set: see Supplementary Table 2). Genes significantly ($P < 0.001$) downregulated are indicated (JQ1, yellow; PD901, blue; PD901/JQ1, black). Genes uniquely or more potently suppressed by PD901/JQ1 are indicated in green. BRD4-bound genes are indicated in red. H3K27Ac and BRD4 were significantly enriched at the promoters of downregulated genes (compared with overall genes bound; 72% versus 46%, Fisher's exact test, $P = 7.9 \times 10^{-3}$). A red box with black square denotes a minimum fourfold displacement of BRD4 in response to PD901/JQ1. H3K27Ac and BRD4 binding overlaps at 69/71 promoters. **c**, Box plot of ssGSEA comparing the enrichment of our Ras signature gene set after treatments. The output shows cooperativity of JQ1 and PD901 in suppressing this Ras signature ($n = 3$, biological replicate). **d**, Gene tracks of H3K27Ac and BRD4 ChIP-seq occupancy at *SOX9* after DMSO or PD901/JQ1 treatment (y axis, reads per million base pairs). **e**, Waterfall plot of MPNSTs from *Nf1/p53/Suz12* mice treated with vehicle control, PD-0325901, JQ1 or a combination of PD-0325901 and JQ1 (10 days). **f**, Immunoblots of H3K27me3 in lung tissue of WT, *Nf1/p53*, *SUZ12*^{+/-} and *Nf1/p53/SUZ12* mice. H3K27me3 levels are disproportionately low in all animals carrying one mutant *SUZ12* allele. **g**, Waterfall plot of MPNSTs from *Nf1/p53* mice comparing vehicle with combination of PD-0325901 and JQ1 ($P = 5.4 \times 10^{-4}$, t -test). **h**, Immunohistochemistry comparing H3K27me3 expression in a human MPNST that is WT for PRC2 genes (P6) and a *SUZ12* null tumour (P7). Arrow depicts positive staining of blood vessel. An additional tumour set is shown in Extended Data Fig. 4d, and results from all primary MPNSTs that were evaluated are shown in Extended Data Table 1.

BRD4 inhibitors alone may not be effective in these tumours, but should be evaluated in combination with MEK inhibitors. However, this work may also have other important clinical implications. EZH2 inhibitors have therapeutic potential in cancers with mutated or overexpressed *EZH2*. Given the prevalence of *SUZ12* and *EED* alterations, it may be prudent to screen for these events in broader clinical trials, as EZH2 inhibitors could exacerbate potential pro-tumorigenic effects of heterozygous alterations. Regardless, these studies reveal a mechanistic link between Ras and PRC2, and provide a scientific rationale for developing a new therapeutic combination for tumours that are currently untreatable.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 December 2013; accepted 6 June 2014.

Published online 13 August 2014.

1. Sauvageau, M. & Sauvageau, G. Polycomb group proteins: multi-faceted regulators of somatic stem cells and cancer. *Cell Stem Cell* **7**, 299–313 (2010).
2. Ernst, T. et al. Inactivating mutations of the histone methyltransferase gene *EZH2* in myeloid disorders. *Nature Genet.* **42**, 722–726 (2010).
3. Zhang, J. et al. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature* **481**, 157–163 (2012).
4. Maertens, O. & Cichowski, K. An expanding role for RAS GTPase activating proteins (RAS GAPs) in cancer. *Adv. Biol. Regul.* **55**, 1–14 (2014).
5. Downward, J. Targeting RAS signalling pathways in cancer therapy. *Nature Rev. Cancer* **3**, 11–22 (2003).
6. Shaw, A. T. et al. Sprouty-2 regulates oncogenic K-ras in lung development and tumorigenesis. *Genes Dev.* **21**, 694–707 (2007).
7. Feldser, D. M. et al. Stage-specific sensitivity to p53 restoration during lung cancer progression. *Nature* **468**, 572–575 (2010).
8. Junttila, M. R. et al. Selective activation of p53-mediated tumour suppression in high-grade tumours. *Nature* **468**, 567–571 (2010).
9. Xu, J. et al. Dominant role of oncogene dosage and absence of tumor suppressor activity in Nras-driven hematopoietic transformation. *Cancer Discov.* **3**, 993–1001 (2013).
10. McLaughlin, S. K. et al. The RasGAP gene, *RASAL2*, is a tumor and metastasis suppressor. *Cancer Cell* **24**, 365–378 (2013).
11. Min, J. et al. An oncogene-tumor suppressor cascade drives metastatic prostate cancer by coordinately activating Ras and nuclear factor- κ B. *Nature Med.* **16**, 286–294 (2010).
12. De Raedt, T. et al. Genomic organization and evolution of the *NF1* microdeletion region. *Genomics* **84**, 346–360 (2004).
13. Lopez-Corra, C. et al. Recombination hotspot in *NF1* microdeletion patients. *Hum. Mol. Genet.* **10**, 1387–1392 (2001).

14. De Raedt, T. *et al.* Elevated risk for MPNST in NF1 microdeletion patients. *Am. J. Hum. Genet.* **72**, 1288–1292 (2003).
15. Mautner, V. F. *et al.* Clinical characterisation of 29 neurofibromatosis type-1 patients with molecularly ascertained 1.4 Mb type-1 NF1 deletions. *J. Med. Genet.* **47**, 623–630 (2010).
16. Beroukhi, R. *et al.* Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc Natl Acad Sci USA* **104**, 20007–20012 (2007).
17. McGillicuddy, L. T. *et al.* Proteasomal and genetic inactivation of the NF1 tumor suppressor in gliomagenesis. *Cancer Cell* **16**, 44–54 (2009).
18. Cichowski, K. *et al.* Mouse models of tumor development in neurofibromatosis type 1. *Science* **286**, 2172–2176 (1999).
19. Legius, E. *et al.* TP53 mutations are frequent in malignant NF1 tumors. *Genes Chromosom. Cancer* **10**, 250–255 (1994).
20. Beert, E. *et al.* Atypical neurofibromas in neurofibromatosis type 1 are premalignant tumors. *Genes Chromosom. Cancer* **50**, 1021–1032 (2011).
21. Pasini, D. *et al.* Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes. *Nucleic Acids Res.* **38**, 4958–4969 (2010).
22. Filippakopoulos, P. *et al.* Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell* **149**, 214–231 (2012).
23. Filippakopoulos, P. *et al.* Selective inhibition of BET bromodomains. *Nature* **468**, 1067–1073 (2010).
24. Widemann, B. C. Current status of sporadic and neurofibromatosis type 1-associated malignant peripheral nerve sheath tumors. *Curr. Oncol. Rep.* **11**, 322–328 (2009).
25. Delmore, J. E. *et al.* BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell* **146**, 904–917 (2011).
26. Puissant, A. *et al.* Targeting MYCN in neuroblastoma by BET bromodomain inhibition. *Cancer Discov.* **3**, 308–323 (2013).
27. Johannessen, C. M. *et al.* TORC1 is essential for NF1-associated malignancies. *Curr. Biol.* **18**, 56–62 (2008).
28. De Raedt, T. *et al.* Exploiting cancer cell vulnerabilities to develop a combination therapy for ras-driven tumors. *Cancer Cell* **20**, 400–413 (2011).
29. Chang, T. *et al.* Sustained MEK inhibition abrogates myeloproliferative disease in NF1 mutant mice. *J. Clin. Invest.* **123**, 335–339 (2013).
30. Jessen, W. J. *et al.* MEK inhibition exhibits efficacy in human and mouse neurofibromatosis tumors. *J. Clin. Invest.* **123**, 340–347 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by the following organizations: The US Department of Defense (W81XWH-11-1-0138), the Ludwig Center at DF/HCC and the Children's Tumor Foundation (K.C.); T.D. was a recipient of the Young Investigator Award of the Children's Tumor Foundation; FWO-Flanders G.0784.10N (E.L.). E.B. was a recipient of an Emmanuel Vanderschueren Fellowship from the Vlaamse Liga tegen Kanker, Association Neurofibromatoses et Recklinghausen, Ligue Française Contre les Neurofibromatoses, Association pour la Recherche sur le Cancer, Comité de Paris de la Ligue Contre le Cancer, the French Clinical Research program (PHRC 2002, P. Wolkenstein) and INSERM (Nf1GeneModif project) (M.V. and E.P.). We thank the Platform of Biological Resources, Assistance Publique Hôpitaux de Paris, Hôpital Henri Mondor, Créteil, France, for providing tissue samples.

Author Contributions T.D. and K.C. conceived and designed the functional studies and mouse experiments. E.L., E.B., M.V., E.P. and A.L. conceived, designed and performed the genomic studies. T.D. performed the cellular, mouse, microarray and ChIP-seq experiments. H.B. coordinated human sample acquisition and analysis. E.L., M.U., V.M., H.K. and M.V. provided human MPNST samples. N.O. performed immunohistological staining. K.H. provided the SUZ12 mice and advice. J.L.H. performed pathological analysis. W.C. assisted in evaluating mouse neurofibromas. J.B. provided compounds and advice. T.D. and K.C. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information All microarray data have been deposited in the Gene Expression Omnibus database under accession number GSE52777. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.C. (kcichowski@rics.bwh.harvard.edu) or E.L. (Eric.Legius@uzleuven.be).

METHODS

Human MPNST samples. All patients were diagnosed as having neurofibromatosis type I. Consent was obtained for tissue acquisition according to Institutional review board approved protocol and local ethical committees.

Cell lines and reagents. ST88-14 (J. Fletcher), 90-8TL (E. Legius), S462 (E. Legius), SNF96.2 (ATCC), U251 (ATCC), A172 (ATCC), T98G (ATCC), MEWO (ATCC), WM3526, Colo-741, RKO (ATCC), H460 (ATCC) and H1355 (ATCC) are human MPNST, GBM, melanoma, colon and lung cancer cell lines. 702T is a mouse *Nf1/p53/Suz12*-deficient MPNST cell line. All cell lines were tested and negative for mycoplasma. Antibodies were obtained from the following sources: H3K27Me3 (9733), H3K27Ac (4353), pERK (XP, 4370) and Vinculin (4670) (Cell Signaling Technologies); SUZ12 (sc-46264) (Santa Cruz Biotechnology); actin (A2066) (Sigma). For ChIP-seq, the following antibodies were used: BRD4 (Bethyl, A301-085A) and H3K27Ac (Abcam ab4729).

Cell counting, IC₅₀ and proliferation curves. All counting experiments were done by haemocytometer. Cells were plated in triplicate for each condition (treatment or genetic alteration). The following day the exact number of cells plated was determined (day 0 value) and cells were treated with conditioned media. The number of cells was determined after 3 (normal growth conditions) or 6 (limiting growth factors) days or precisely after three population doublings to determine values of IC₅₀. Fold growth or log₂ of fold of the number of cells was determined versus day 0. When cell death was induced (compared with the number of cells plated on day 0), data were plotted as log₂ of fold cell number (negative values indicate cell death). IC₅₀ values and 95% confidence intervals were determined using GraphPad Prism 5 software.

Array comparative genomic hybridization. Array comparative genomic hybridization experiments (Agilent 244K and 400K) were performed as previously described²⁹. Array comparative genomic hybridization data (Agilent 244K) from 21 independent, high-grade MPNSTs (HM1, 2, 3, 5–8, 10, 12, 13, 15, 17, 19, 20, 22–26, 28, and 30) were used as published previously²⁰. Digestion, labelling and hybridization of tumour and patient- or gender-matched blood DNA were performed according to the manufacturer's instructions (Agilent).

Sequencing. The coding sequences of 14 H3K27me3-related genes were analysed (*AEBP2*, *EED*, *EZH2*, *EZH1*, *HDAC2*, *JARID2*, *PCL1*, *PCL2*, *PCL3*, *RBBP4*, *RBBP7*, *SIRT1*, *SUZ12* and *UTX*) using a targeted next generation sequencing (NGS) approach. Experiments were performed on the NGS platform of the Cochin Hospital, Paris (Assistance Publique, Hôpitaux de Paris, France). Briefly, the custom primers panel targeting the 14 genes (coding exons and IVS boundaries) was designed using the AmpliSeq Designer (Life Technologies). NGS library preparation used the Ion AmpliSeq Library Kit 2.0 according to the manufacturer's instructions (Ion AmpliSeq Library Preparation, publication part number MAN0006735, revision 5.0, July 2013, Life Technologies). Amplified libraries were purified using Agencourt AMPure XP beads (Beckman Coulter). Prior to library pooling and sequencing sample preparation, amplified libraries were validated and quantified using the 2100 Bioanalyzer microfluidic platform (Agilent Technologies). Emulsion PCR used the Ion OneTouch Instrument. Enrichment of the template-positive Ion OneTouch 200 ion sphere particle PCR used the Ion OneTouch ES (Life Technologies). The template-positive ion sphere particles were loaded on Ion 318 chips and sequenced with an Ion PGM sequencer (Life Technologies). Sequence alignment and extraction of single nucleotide polymorphisms and short insertions/deletions were performed using the Variant Caller plugin on the Ion Torrent Browser, and DNA sequences visualized using the Integrated Genomics Viewer (version 2.3) from the Broad Institute. NextGENe software (version 2.3.3, Softgenetics), was also used for sequence alignment, extraction of single nucleotide polymorphisms and short indels, and their visualization.

Real-time PCR. Cell were scraped and dissolved in Trizol reagent (Invitrogen). RNA was treated with DNaseI (Roche) and reverse transcribed using the qScript Reverse transcriptase kit (Quanta). Real-time PCR analysis used the PerfeCTa SYBR Green kit (Quanta) for *BRD4* (5'-CATGGACATGAGCACAATCA-3' 5'-TCATGGTCAGGAGGGTTGTA-3').

RNA interference and complementary DNA expression. Non-targeting short interfering RNAs (siRNAs) (D-001810-10: 5'-UGGUUUACAUGUCGACUAA-3'; 5'-UGGUUUACAUGUUGUGUGA-3'; 5'-UGGUUUACAUGUUUUCUGA-3'; 5'-UGGUUUACAUGUUUUCUUA-3') and siRNAs against *BRD4* (ON-TARGET smartpool siRNA L-004937-00: 5'-UGGUUUACAUGUCGACUAA-3'; 5'-UGGUUUACAUGUUGUGUGA-3'; 5'-UGGUUUACAUGUUUUCUGA-3'; 5'-UGGUUUACAUGUUUUCUUA-3') were purchased from Dharmacon. siRNAs were transfected with lipofectamine RNAiMAX from Invitrogen. Lentiviral plKO vectors were obtained from the RNAi consortium of the Broad Institute. shSUZ12 S1 (NM_015355.1-2076s1c1, target sequence 5'-GCTGACAATCAAATGAATCAT-3'), shSUZ12 S2 (NM_015355.1-668s1c1, target sequence 5'-GCTTACGTTTACTGGTTTCIT-3') and shSUZ12 S3 (NM_015355.1-501s1c1, target sequence 5'-CGGAATCTCATAGC ACCAATA-3') were used to target *SUZ12*. Human *SUZ12* (Open Biosystems) was

cloned into a pLenti CMV/TO Puro vector. Lentiviral infections were performed as previously described²⁸.

Expression micro-array and gene set enrichment analysis. The MPNST cell line 90-8TL was infected with LACZ control or SUZ12 and cultured for 14 days. The LACZ expressing 90-8TL cell line was also treated with 100 nM JQ1, PD0325901 2 μ M, a combination of PD0325901 (2 μ M) and JQ1 (100 nM) or vehicle (DMSO) for 24 h. Additionally SUZ12 was knocked down in the U251 cell line by shRNA. All experiments were performed in triplicate. RNA was extracted using Trizol reagent (Invitrogen) and purified using Qiagen RNeasy extraction kit (according to the manufacturer's instructions) and hybridized to the Affymetrix Human Gene 1.0 ST array by the Partners HealthCare Center for Personalized Genetic Medicine core. Analyses used BRB-ArrayTools developed by R. Simon and the BRB-ArrayTools development team. Thresholds were set at $P < 0.001$. Supplementary Table 3 lists all the genes significantly changed ($P < 0.001$) compared with control. We first made a *SUZ12* gene list of genes that were downregulated at least twofold ($P < 0.001$) in *SUZ12* reconstituted cells (Supplementary Table 1). GSEA³¹ and ssGSEA used the Broad Institute interface (www.broadinstitute.com/gsea/index.jsp). Several gene lists upregulated by Ras activation were significantly enriched in our experiments (BILD_HRAS_ONCOGENIC_SIGNATURE, KRAS.300_UP.V1_UP, KRAS.600_UP.V1_UP, KRAS.600.LUNG.BREAST_UP.V1_UP, KRAS.BREAST_UP.V1_UP). To identify more specifically a subset of direct suppressive effects on the Ras pathway, we first compiled these five significant gene lists and selected those genes that were significantly repressed by JQ1, PD-901 or combined JQ1/PD901 in MPNSTs (Supplementary Table 2) (class comparison of all treatments at $P < 0.001$).

ChIP-seq analysis. BRD4, H3K27Me3 and H3K27Ac were ChIPed following the Agilent Mammalian ChIP-on-chip protocol. 90-8TL cells were grown and treated under the same conditions as performed for microarray studies. Cells were cross-linked for 15 min at room temperature by the addition of one-tenth of the volume of 1% formaldehyde solution to the growth media. Pull down was performed with the indicated antibodies and DNA was submitted for sequencing to the CCCB core of the Dana-Farber Cancer Institute.

Alignment used the BWA package. Mapped tags were analysed using a statistical method that fitted a Poisson distribution to the data and calculated a global false discovery rate for peak detection. Peak calling used the HOMER software package with the false discovery rate set to 0.001. Peak calling in this setting is differential because the algorithm requires that the number of tags mapped to an identified peak region is fourfold higher in the test sample than the input sample. A background filter was applied by analysing a 10 kilobase region surrounding each peak. Peaks were removed unless tag density within each peak was at least fourfold higher than in the surrounding region. Peaks were only called in regions of the genome that were not masked by standard repeat element filters. Although repeat areas were removed, some regions of the genome still had low mappability due to undetected repeats or genome duplications. Accordingly, a final filter removed peaks in such poorly mappable regions. Differential peaks (comparing binding in different treatment groups) were also determined by the HOMER package.

Mouse models and pathology. Animal procedures were approved by the Center for Animal and Comparative Medicine in Harvard Medical School in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals and the Animal Welfare Act. *c57bl/6 Nf1^{+/-}* (ref. 32) and *Suz12^{+/-}* (ref. 33) were crossed to generate double heterozygous mice. These mice were intercrossed further to generate *Nf1^{+/-}; Suz12^{+/-}* mice with both genes in a *cis* configuration. Similarly *Nf1^{+/-}; Suz12^{+/-}* mice were crossed to *p53^{+/-}* mice to generate NPS mice having *Nf1*, *Suz12* and *p53* mutations in *cis* on chromosome 11. For our survival studies we used 15 animals in control groups and 30 animals in test groups. Relevant tumour types were evaluated by a human pathologist (J.H.). The pathologist was blinded to the group allocation.

Malignant peripheral nerve sheath tumour. The tumours are composed of fascicles of spindle cells with elongated, wavy nuclei, coarse chromatin and pale, eosinophilic cytoplasm. Scattered tumour cells show marked pleomorphism, including multinucleated forms. There are frequent mitotic figures.

High-grade glioma. The brain is infiltrated by a markedly hypercellular tumour composed of rounded cells with hyperchromatic to vesicular nuclei and variably prominent nucleoli, including occasional large and multinucleated cells.

Olfactory neuroblastoma. Tumours are composed of primitive polygonal cells with vesicular chromatin, prominent, often multiple nucleoli, amphophilic cytoplasm and frequent mitotic figures. The tumour cells are arranged in sheets and form numerous rosettes, and there is multifocal necrosis.

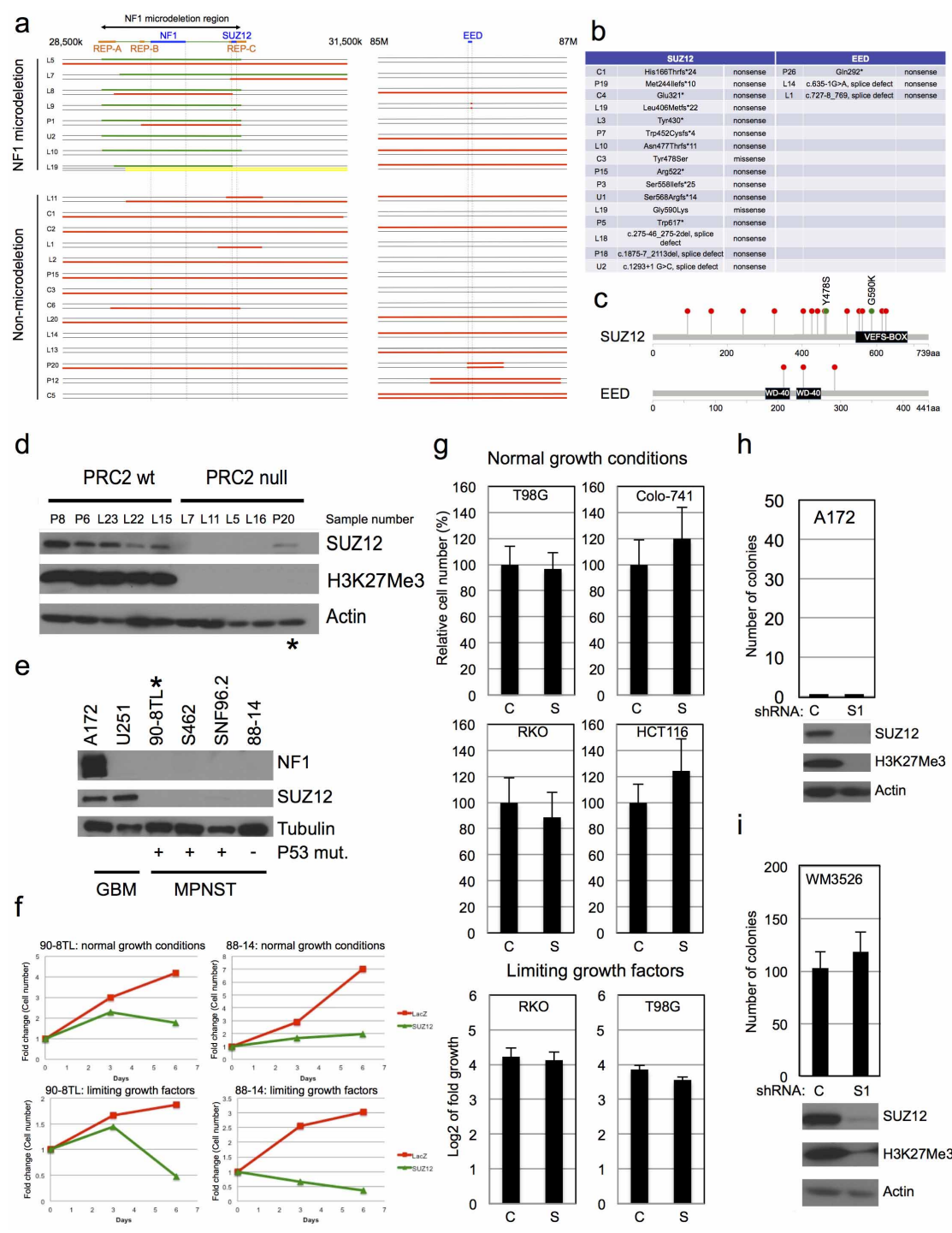
Melanocytic nevus. Heavily black-pigmented lesion (consistent with melanin). The lesional cells are uniform with small nuclei and no nuclear atypia or mitotic activity. The lesion is well circumscribed and does not infiltrate into adjacent tissues.

Intestinal adenoma. These dysplastic polyps (adenomas) show a tubular architecture and are composed of columnar cells with enlarged and elongated, hyperchromatic nuclei with prominent nucleoli and increased mitotic activity.

Drug treatment, dosing schedule and tumour volume measurements. Only mice that developed an MPNST that was readily measurable by Vernier callipers were included in the study. As these tumours develop spontaneously over time, and to ensure adequate randomization, mice were sequentially enrolled per treatment arm, starting with the combination treatment arm. The investigator was not blinded. To have sufficient power (80%) to detect a minimum of 50% tumour shrinkage versus day 0 ($P = 0.05$), we enrolled six animals per treatment group. Mice were enrolled in the study when tumour size reached 300–700 mm³. Tumour size was measured every 2–3 days by Vernier callipers. Tumour volume was calculated using the standard formula $L \times W^2 \times 52$. Tumour volume and log₂ of fold growth versus day 0 were calculated and plotted on a graph. JQ1 (45 mg/kg) was administered intraperitoneally daily in a 10% (2-hydroxypropyl)- β -cyclodextrin solution (C0926, Sigma). PD-0325901 was administered by oral gavage daily at 1.5 mg/kg (vehicle (0.5% (w/v) methylcellulose solution with 0.2% (v/v) polysorbate 80 (Tween 80))). Compounds given in combination were administered sequentially.

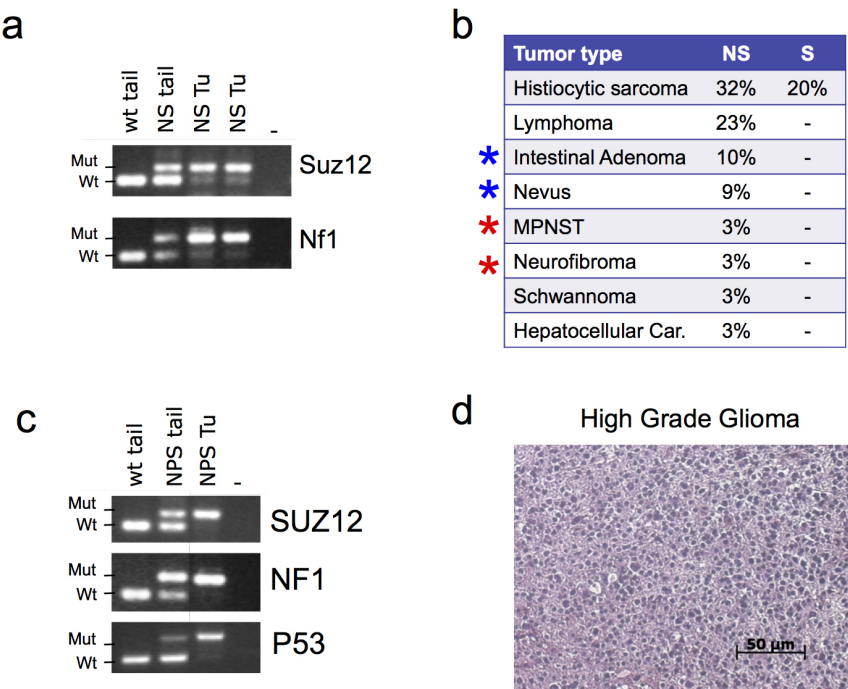
Statistics. Unless otherwise stated, all statistical analysis used SYSTAT 12 software. For each data set, basic statistical values (mean and s.d.) were calculated and normality determined (Shapiro–Wilk normality test); all data sets were normally distributed. All fold change tumour volumes for the different drug combinations were compared by Student's *t*-test (unequal variance). Survival analysis used the Mantel–Cox test. The statistical tests used are mentioned throughout the text.

31. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
32. Jacks, T. *et al.* Tumour predisposition in mice heterozygous for a targeted mutation in Nf1. *Nature Genet.* **7**, 353–361 (1994).
33. Pasini, D., Bracken, A. P., Jensen, M. R., Lazzerini Denchi, E. & Helin, K. Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *EMBO J.* **23**, 4061–4071 (2004).



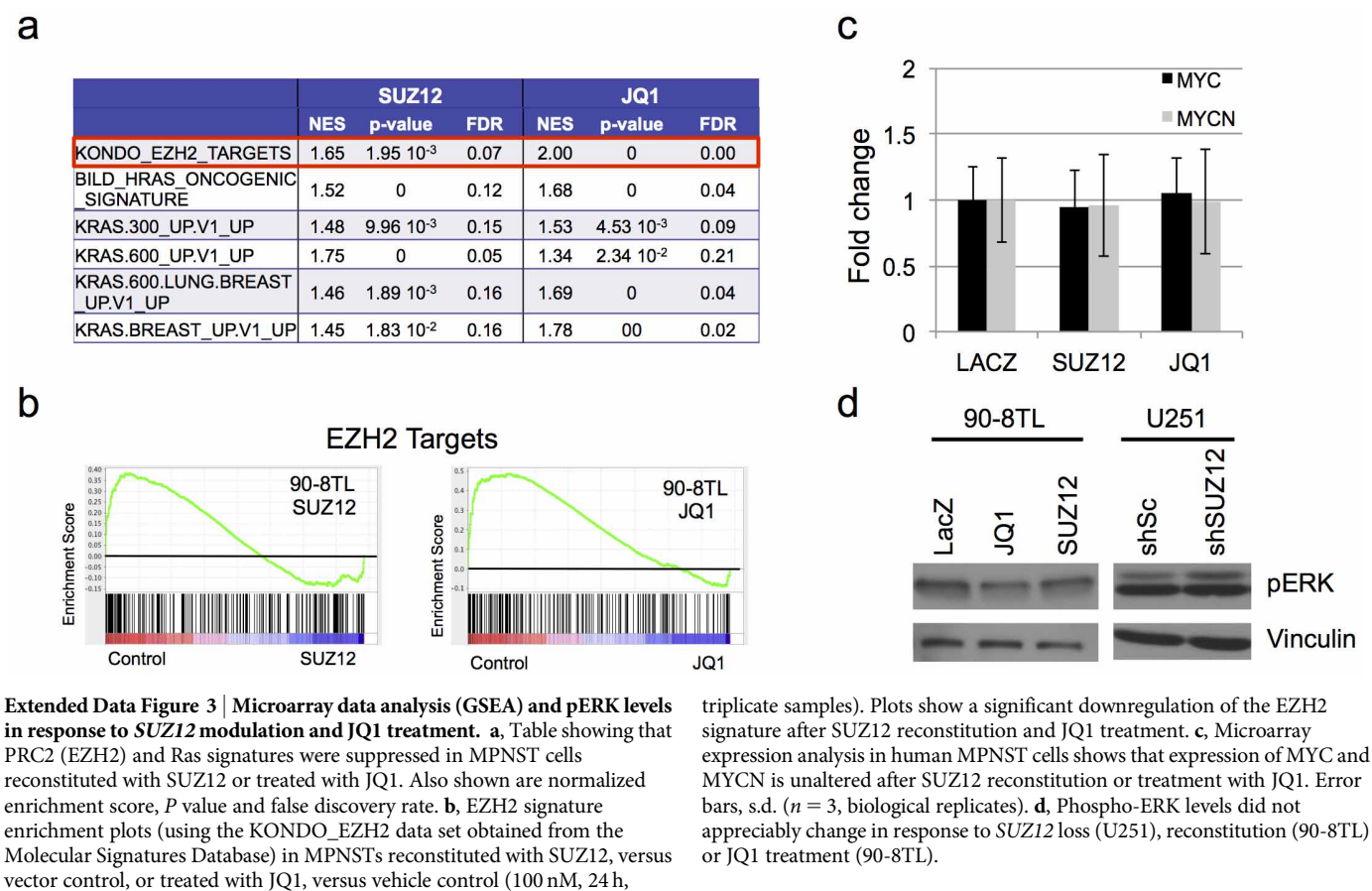
Extended Data Figure 1 | Mutational data in human MPNSTs and further biological analysis of SUZ12 loss. **a**, Schematic overview of the deletions in the *NF1*, *SUZ12* and *EED* regions observed in human MPNSTs (green, germline deletion; red, somatic deletion; yellow, duplication). **b**, List of the amino-acid changes or deletions found in *SUZ12* and *EED* in human MPNSTs. **c**, Schematic representation of the location of *SUZ12* and *EED* mutations (red, truncating mutation; green, missense mutation with amino-acid change noted). **d**, Immunoblots of lysates from primary human MPNSTs. Tumours with homozygous inactivating mutations in one of the PRC2 components (*SUZ12* or *EED*) show complete loss of H3K27Me3. *Homozygous inactivation of *EED*. **e**, Immunoblots comparing NF1, *SUZ12* and H3K27me3 expression of four human MPNST cell lines. *Cell line derived from an MPNST of a patient with an *NF1* microdeletion. The human GBM cell lines A172 and U251 were used as a

control. Corresponding *NF1* mutations are reported in Extended Data Table 1 (S462, L2; 90-8TL, L3) and elsewhere in these *NF1*-deficient lines. p53 mutations, when known, are denoted and reported elsewhere²⁰. **f**, Proliferation curves used to derive bar graphs shown in Fig. 1d (red, LacZ control; green, *SUZ12* reconstituted). **g**, Relative proliferation of several *SUZ12* WT cell lines: colon (RKO, colo741, HCT-116) or GBM (T98G), after introduction of a control or *SUZ12* lentivirus. None of these cell lines exhibited a significant decrease in proliferation under normal growth conditions or cell death in limiting growth factors, in contrast to *SUZ12*-deficient cells shown in Fig. 1d. **h**, Effects of sh*SUZ12* (S1) on colony formation (top) and *SUZ12* expression (bottom) in A172 GBM cells, which are *NF1* WT (see Extended Data Fig. 1e). **i**, Effects of sh*SUZ12* (S1) on *SUZ12* expression and colony formation in WM3526 melanoma cells, which are *NF1* WT. Error bars, s.d. ($n = 3$, biological replicates).



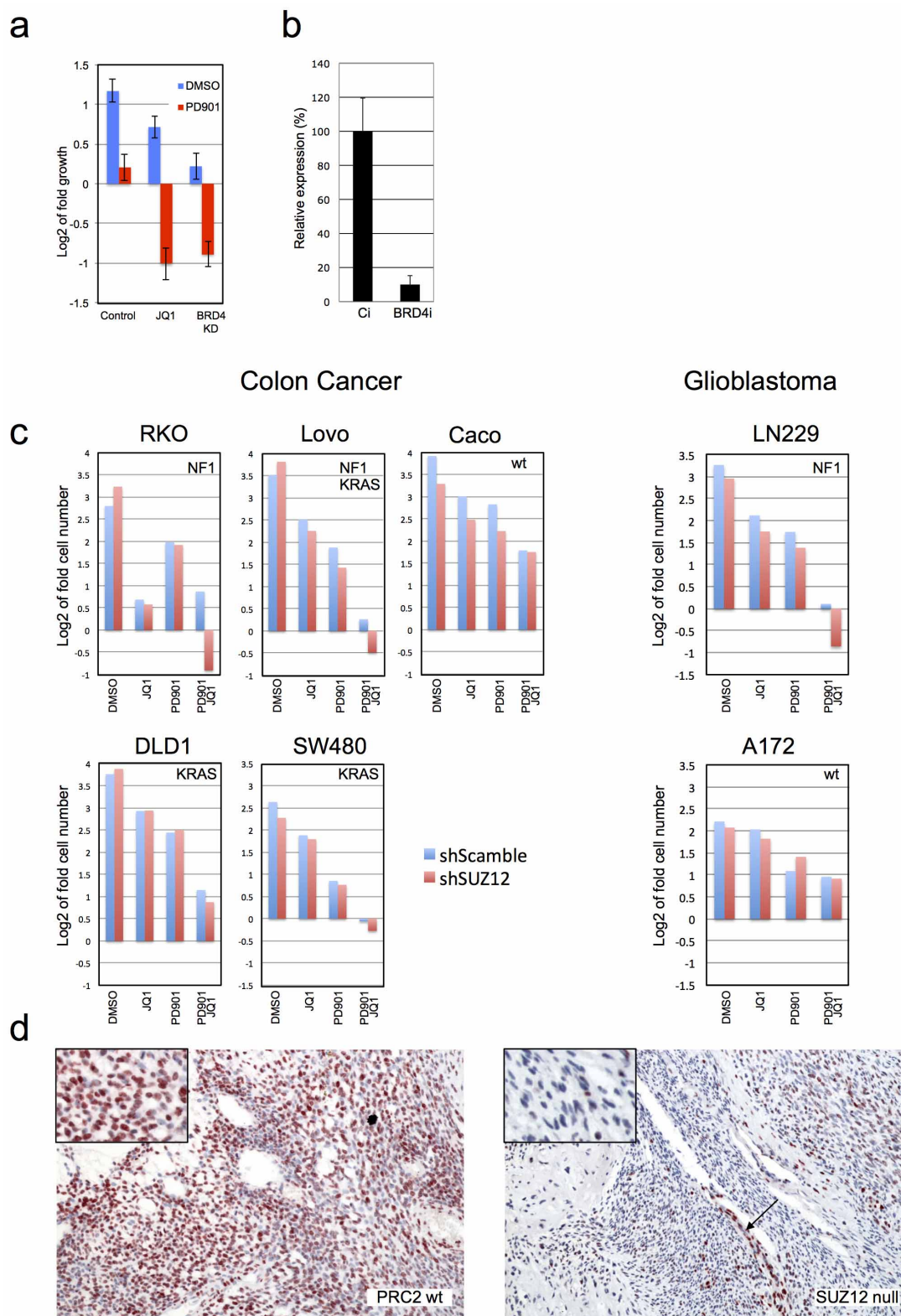
Extended Data Figure 2 | *Suz12* and *Nf1* mutations cooperate to promote widespread tumour development in mice. **a**, Semi-quantitative PCR showing loss of the WT *Suz12* and *Nf1* allele in *Nf1/Suz12* mouse tumours. **b**, Table listing the tumours observed in *Nf1*^{+/-}; *Suz12*^{+/-} and *Suz12*^{+/-} mice. Tumour types denoted with a red asterisk occur in patients with NF1 and an increased frequency in patients with NF1 microdeletions. Tumour types

denoted with a blue asterisk represent spontaneous tumour types/lesions that have been shown to harbour NF1 mutations humans. **c**, Semi-quantitative PCR showing loss of the WT *Suz12*, *Nf1* and *p53* allele in *Nf1/p53/Suz12* mouse MPNST. **d**, Haematoxylin and eosin staining of a GBM from an *Nf1/p53/Suz12* mouse.



Extended Data Figure 3 | Microarray data analysis (GSEA) and pERK levels in response to *SUZ12* modulation and JQ1 treatment. **a**, Table showing that PRC2 (EZH2) and Ras signatures were suppressed in MPNST cells reconstituted with SUZ12 or treated with JQ1. Also shown are normalized enrichment score, *P* value and false discovery rate. **b**, EZH2 signature enrichment plots (using the KONDO_EZH2 data set obtained from the Molecular Signatures Database) in MPNSTs reconstituted with SUZ12, versus vector control, or treated with JQ1, versus vehicle control (100 nM, 24 h,

triplicate samples). Plots show a significant downregulation of the EZH2 signature after SUZ12 reconstitution and JQ1 treatment. **c**, Microarray expression analysis in human MPNST cells shows that expression of MYC and MYCN is unaltered after SUZ12 reconstitution or treatment with JQ1. Error bars, s.d. ($n = 3$, biological replicates). **d**, Phospho-ERK levels did not appreciably change in response to *SUZ12* loss (U251), reconstitution (90-8TL) or JQ1 treatment (90-8TL).



Extended Data Figure 4 | Consequences of SUZ12 loss on drug sensitivity and H3K27me3 levels. **a**, Graph depicting log₂ fold change in MPNST cell number (90-8TL) after 3 days in response to drugs and siRNA sequences indicated. PD-0325901 also cooperates with genetic BRD4 ablation to kill cells. **b**, Relative expression of BRD4 as determined by real-time PCR after exposure to BRD4-specific siRNAs (90-8TL). **c**, SUZ12 ablation using shSUZ12 S1 conferred increased sensitivity to combined PD-901/JQ1 in *NF1*-deficient but not *NF1* WT colon cancer and GBM cell lines. Log₂ fold change in cell number was determined. RKO, DLD1 and Caco cell lines were treated with 1 μ M JQ1

and 1 μ M PD901 each time alone or in combination. The Lovo cell line was treated with 500 nM JQ1 and 250 M PD901; SW480 was treated with 250 nM PD901 and 1 μ M JQ1; LN229 and A172 cell lines were treated with 500 nM JQ1 and 1 μ M PD901. **d**, Immunohistochemistry comparing H3K27me3 expression in an additional human MPNST that was WT for *SUZ12* and other PRC2 genes, and one that was null for *SUZ12*. Arrow depicts positive staining of blood vessel. Staining of additional tumours is shown in Extended Data Table 1. Error bars, s.d. ($n = 3$, biological replicates).

Extended Data Table 1 | Combined deletion and mutational data for tumours listed in Fig. 1

	SAMPLE	NF1		SUZ12		EED		H3K27Me3 IHC
		germline	somatic	first hit	second hit	somatic 1	somatic 2	
NFI	L10M12	type 1 MD	not detected, mutation NE	germline type 1 MD	somatic c.1400del, p.(Gln677Met)*11	chr11:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	-	NE
	L16 (NM20)	type 1 MD	not detected, mutation NE	germline type 1 MD	somatic c.1358del, p.(Gln450Arg)	-	-	NE
	L19 (NM24)	type 1 MD	not detected, mutation NE	germline type 1 MD	somatic c.1358del, p.(Gln450Arg)	-	-	Positive
	L3 (NM21) + 40,471	type 1 MD	not detected, mutation NE	germline type 1 MD	somatic c.1358del, p.(Gln450Arg)	chr11:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	-	NE
	L4 (NM5)	type 1 MD	chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	germline type 1 MD	somatic c.1358del, p.(Gln450Arg)	-	-	NE
	L7 (NM7)	type 1 MD	chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	germline type 1 MD	somatic c.1358del, p.(Gln450Arg)	-	-	NE
	L8 (NM8)	type 1 MD	not detected, mutation NE	germline type 1 MD	somatic c.1358del, p.(Gln450Arg)	-	-	Loss
	L9 (NM22)	type 1 MD	not detected, mutation NE	germline type 1 MD	somatic c.1358del, p.(Gln450Arg)	-	-	Loss
	P1	type 1 MD	type 1 MD	germline type 1 MD	somatic c.1358del, p.(Gln450Arg)	-	-	Loss
	P10	chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	c.679C>T, p.(Gln226*)	germline chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	-	-	-	Loss
E	P11	chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	c.499, 302del, p.(Gln450Arg)	germline chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	-	-	-	NE
	P12	chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	c.755G>T, p.(Gln250*)	germline chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	-	-	-	NE
	P13	type 1 MD	c.244del, p.(Gln81*)	germline type 1 MD	somatic c.1358del, p.(Gln450Arg)	-	-	NE
	P2	chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	c.672G>A, splice defect	germline chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	-	-	-	NE
	P16	chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	c.432T>A, p.(Gln143*)	germline chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	-	-	-	NE
	P5	chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	c.290T>C, p.(Gln96Arg)	germline chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	-	-	-	NE
	P7	type 1 MD	c.138del, p.(Gln45*)	germline type 1 MD	somatic c.1358del, p.(Gln450Arg)	-	-	Loss
	U1	type 1 MD	NE	germline type 1 MD	somatic c.1358del, p.(Gln450Arg)	-	-	NE
	U2	type 1 MD	NE	germline type 1 MD	somatic c.1358del, p.(Gln450Arg)	-	-	NE
	C1	c.488del, p.(Gln162Arg)*10	chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic c.1358del, p.(Gln450Arg)	chr11:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	not detected, mutation NE	NE
N	C2	c.2012C>G, splice defect	chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic c.1358del, p.(Gln450Arg)	chr11:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	not detected, mutation NE	NE
	C3	chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic c.1358del, p.(Gln450Arg)	chr11:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	not detected, mutation NE	NE
	C4	c.177del, p.(Gln59Arg)*10	c.634G>A, splice defect	somatic chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic c.1358del, p.(Gln450Arg)	chr11:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	not detected, mutation NE	NE
	C5	c.372G>T, p.(Gln124*)	chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic c.1358del, p.(Gln450Arg)	chr11:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	not detected, mutation NE	NE
	C7	c.304G>T, p.(Gln101*)	chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic c.1358del, p.(Gln450Arg)	chr11:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	not detected, mutation NE	NE
	L1 (NM1)	c.493G>T, p.(Gln164*)	chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic c.1358del, p.(Gln450Arg)	chr11:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	not detected, mutation NE	NE
	L3 (NM21)	c.138G>T, p.(Gln46*)	chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic c.1358del, p.(Gln450Arg)	chr11:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	not detected, mutation NE	NE
	L3 (NM21)	Not detected, mutation NE	not detected, mutation NE	somatic chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic c.1358del, p.(Gln450Arg)	chr11:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	not detected, mutation NE	Positive
	L4 (NM17)	c.589C>T, p.(Gln196*)	chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic c.1358del, p.(Gln450Arg)	chr11:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	not detected, mutation NE	Loss
	L5 (NM19)	c.1732G>A, splice defect	chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic c.1358del, p.(Gln450Arg)	chr11:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	not detected, mutation NE	Positive
M	L7 (NM22)	c.133G>T, splice defect	chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic c.1358del, p.(Gln450Arg)	chr11:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	not detected, mutation NE	NE
	L18 (NM22) + 88.3	c.690T>C, p.(Gln231Arg)	not detected, mutation NE	somatic chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic c.1358del, p.(Gln450Arg)	chr11:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	not detected, mutation NE	NE
	L2 (NM22) + 54.2	c.670G>A, splice defect	chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic c.1358del, p.(Gln450Arg)	chr11:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	not detected, mutation NE	NE
	L20 (NM25)	Not detected, mutation NE	not detected, mutation NE	somatic chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic c.1358del, p.(Gln450Arg)	chr11:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	not detected, mutation NE	Loss
	L21 (NM26)	-	not detected, mutation NE	somatic chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic c.1358del, p.(Gln450Arg)	chr11:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	not detected, mutation NE	Positive
	L22 (NM28)	c.780G>T, p.(Gln260*)	not detected, mutation NE	somatic chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic c.1358del, p.(Gln450Arg)	chr11:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	not detected, mutation NE	Positive
	L23 (NM28)	Not detected, mutation NE	not detected, mutation NE	somatic chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic c.1358del, p.(Gln450Arg)	chr11:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	not detected, mutation NE	Positive
	L2 (NM26)	Not detected, mutation NE	not detected, mutation NE	somatic chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic c.1358del, p.(Gln450Arg)	chr11:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	not detected, mutation NE	Loss
	P12	c.589C>T, p.(Gln196*)	c.254G>T, p.(Gln85Arg)	somatic chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic c.1358del, p.(Gln450Arg)	chr11:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	not detected, mutation NE	Loss
	P13	c.188G>A, p.(Gln63Arg)*4	-	somatic chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic c.1358del, p.(Gln450Arg)	chr11:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	not detected, mutation NE	NE
O	P14	c.188G>A, p.(Gln63Arg)*4	-	somatic chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic c.1358del, p.(Gln450Arg)	chr11:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	not detected, mutation NE	Loss
	P15	c.188G>A, p.(Gln63Arg)*4	-	somatic chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic c.1358del, p.(Gln450Arg)	chr11:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	not detected, mutation NE	NE
	P17	c.138G>T, p.(Gln46*)	chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic c.1358del, p.(Gln450Arg)	chr11:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	not detected, mutation NE	NE
	P19	c.138G>T, p.(Gln46*)	chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic c.1358del, p.(Gln450Arg)	chr11:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	not detected, mutation NE	NE
	P3	chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic c.1358del, p.(Gln450Arg)	chr11:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	not detected, mutation NE	Loss
	P4	c.30G>T, p.(Gln10*)	c.725G>A, p.(Gln241Arg)*3	somatic chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic c.1358del, p.(Gln450Arg)	chr11:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	not detected, mutation NE	Loss
	P6	c.295C>T, p.(Gln98*)	c.678G>A, p.(Gln227Arg)*5	somatic chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic c.1358del, p.(Gln450Arg)	chr11:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	not detected, mutation NE	Positive
	P8	c.239T>C, p.(Gln79Arg)	-	somatic chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic c.1358del, p.(Gln450Arg)	chr11:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	not detected, mutation NE	Positive
	P9	c.138G>T, p.(Gln46*)	-	somatic chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic c.1358del, p.(Gln450Arg)	chr11:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	not detected, mutation NE	Positive
	P10	c.138G>T, p.(Gln46*)	-	somatic chr17:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	somatic c.1358del, p.(Gln450Arg)	chr11:hg19:g20,371,409_32,341,496, (134,927,035_134,927,114)del	not detected, mutation NE	Positive

Mutations and resulting amino-acid changes in *NF1*, *SUZ12* and *EED* are shown. H3K27me3 staining results, when performed, are indicated and tumours are classified as strongly positive, negative (loss) or partial loss (ploss). The majority of MPNSTs harbour defects in the p53 pathway.

Mechanism of Dis3l2 substrate recognition in the Lin28–let–7 pathway

Christopher R. Faehnle^{1,2*}, Jack Walleshauser^{1,2,3*} & Leemor Joshua-Tor^{1,2,3,4}

The pluripotency factor Lin28 inhibits the biogenesis of the let-7 family of mammalian microRNAs^{1–4}. Lin28 is highly expressed in embryonic stem cells and has a fundamental role in regulation of development⁵, glucose metabolism⁶ and tissue regeneration⁷. Overexpression of Lin28 is correlated with the onset of numerous cancers⁸, whereas let-7, a tumour suppressor, silences several human oncogenes⁵. Lin28 binds to precursor let-7 (pre-let-7) hairpins⁹, triggering the 3' oligo-uridylation activity of TUT4 and TUT7 (refs 10–12). The oligoU tail added to pre-let-7 serves as a decay signal, as it is rapidly degraded by Dis3l2 (refs 13, 14), a homologue of the catalytic subunit of the RNA exosome. The molecular basis of Lin28-mediated recruitment of TUT4 and TUT7 to pre-let-7 and its subsequent degradation by Dis3l2 is largely unknown. To examine the mechanism of Dis3l2 substrate recognition we determined the structure of mouse Dis3l2 in complex with an oligoU RNA to mimic the uridylated tail of pre-let-7. Three RNA-binding domains form an open funnel on one face of the catalytic domain that allows RNA to navigate a path to the active site different from that of its exosome counterpart. The resulting path reveals an extensive network of uracil-specific interactions spanning the first 12 nucleotides of an oligoU-tailed RNA. We identify three U-specificity zones that explain how Dis3l2 recognizes, binds and processes uridylated pre-let-7 in the final step of the Lin28–let-7 pathway.

In embryonic stem cells, Dis3l2 functions as the effector nuclease responsible for degrading uridylated pre-let-7 miRNAs in the Lin28–let-7 pathway^{13,14}. Dis3l2 belongs to the RNase II/R 3'–5' exonuclease superfamily, which includes the catalytic subunit of the RNA exosome in yeast (Rrp44)^{15,16} and humans (Dis3 and Dis3l1)¹⁷. Inactivation of Dis3l2 is associated with aneuploidy and mitotic abnormalities, whereas its overexpression suppresses cancer cell growth¹⁸. Genetic disruption of *DIS3L2* is the primary cause of Perlman's syndrome, a congenital disorder leading to fetal overgrowth and an increased susceptibility to Wilms' tumour development¹⁸. Subsequent studies have shown that Wilms' tumours, a common paediatric kidney cancer, overexpress Lin28 (ref. 8), underscoring the role of miRNA regulation in kidney tumorigenesis. Outside of miRNA regulation, Dis3l2 mediates 3'–5' mRNA decay in an alternative mRNA decay pathway that is independent of the exosome^{19,20}. This is particularly intriguing in light of studies showing widespread uridylation of mammalian mRNAs^{21,22}, suggesting that RNA uridylation may be a common RNA degradation signal. Here we present the structure of mouse Dis3l2 in complex with an oligoU RNA and provide an explanation for Dis3l2's high specificity towards uridylated RNA substrates.

It was shown that Dis3l2 is a processive 3'–5' hydrolytic exonuclease that preferentially degrades uridylated pre-let-7 and other uridylated miRNAs^{13,14}. Related studies showed that Dis3l2 processed structured RNA more efficiently than its exosome counterparts (human Dis3 and yeast Rrp44), with no reported sequence preference²⁰. However, the fission yeast Dis3l2 orthologue was shown to exhibit a preference for oligoU¹⁹. To clarify the extent of mammalian Dis3l2's substrate preference, we quantitatively measured the specific activity of mouse Dis3l2

for single-stranded RNA substrates (Fig. 1a and Extended Data Fig. 1a, b). Dis3l2 processed U₁₅ ~9-fold and ~40-fold more efficiently than C₁₅ and A₁₅, respectively, which are both relatively poor substrates (Fig. 1a). Dis3l2 failed to process deoxyU₁₅, thus is specific for RNA (Extended Data Fig. 1a). Next, we measured binding affinity, where Dis3l2 exhibited more than 200-fold tighter binding to U₁₅ compared to A₁₅ and C₁₅ (Fig. 1b and Extended Data Fig. 1c, d). Notably, the binding constant for deoxyU₁₅ was comparable to U₁₅, which further illustrates its preference for uracil (Fig. 1b). We extended our analysis to the known biological substrate, pre-let-7. Dis3l2 degraded pre-let-7-U₁₅ more efficiently than the unmodified pre-let-7 hairpin and had a tenfold higher affinity for pre-let-7-U₁₅ over pre-let-7 (Fig. 1a, b). Taken together, both enhanced activity and tighter binding have a role in Dis3l2's notable preference for oligoU.

To understand the substrate preference of Dis3l2 we determined the crystal structure of mouse Dis3l2 in complex with an oligoU RNA substrate (Fig. 1c and Extended Data Table 1). Full-length Dis3l2 failed to crystallize. We therefore crystallized a truncated form of Dis3l2 encompassing residues 37–857 with two flexible loops removed (residues 148–169 and 194–221). The construct also contained an inactivating mutation (D389N) in the active site to trap oligoU (Extended Data Fig. 2). The truncated form of Dis3l2 with an intact active site had comparable substrate preference and activity as the full-length enzyme (Extended Data Fig. 3a). Unambiguous electron density accounted for the modelling of 14 nucleotides of oligoU (U₁₄) (Extended Data Fig. 3b, see full description in Methods). Dis3l2 is composed of two cold shock domains at its amino terminus (CSD1 and CSD2), which are intimately associated with each other (Extended Data Fig. 4a, b) and form a lobe on the 'top' side of the catalytic RNB domain, while the carboxy-terminal S1 domain sits opposite the CSD lobe (Fig. 1c, d and Extended Data Fig. 4c). The CSD lobe and S1 form an open funnel on 'top' of the RNB, where RNA substrates gain entry to the enzyme (Fig. 1c–e). The entire length of U₁₄ stretches from the opening of the funnel at the top and threads through the RNB to the active site. The 5'-half of U₁₄ (U1–U7) wraps around the mouth of the funnel, primarily interacting with loops protruding from the CSD lobe and S1 (Fig. 1d and Extended Data Fig. 4d). This is followed by U8, which resides in a pocket formed by S1 and RNB at the stem of the funnel (Extended Data Fig. 4c). The final 6 nucleotides, U9–U14, are stacked within the RNB core (Fig. 1c–e). The 3'-end nucleotides, U13–U14, are situated in the active site in the 'middle' of the RNB domain (Fig. 1c), where there is a solvent-accessible escape 'hole' through the side of the enzyme for nucleotide products to exit. We measured specific activity and binding affinity for oligoU substrates of increasing length (9–15 nucleotides) and determined that the optimal oligoU substrate (U₁₅) spans the entire length of the enzyme (Extended Data Fig. 5). The Dis3l2–U₁₄ structure, therefore, represents the optimal length oligoU tail required to recruit Dis3l2 and elicit the enhanced degradation of pre-let-7 (refs 13, 14) and is consistent with the length of oligoU tail added to pre-let-7 by TUT4 and TUT7 (refs 10–12).

¹W. M. Keck Structural Biology Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA. ²Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA.

³Watson School of Biological Science, Cold Spring Harbor, 1 Bungtown Road, New York 11724, USA. ⁴Howard Hughes Medical Institute, Cold Spring Harbor, 1 Bungtown Road, New York 11724, USA.

*These authors contributed equally to this work.

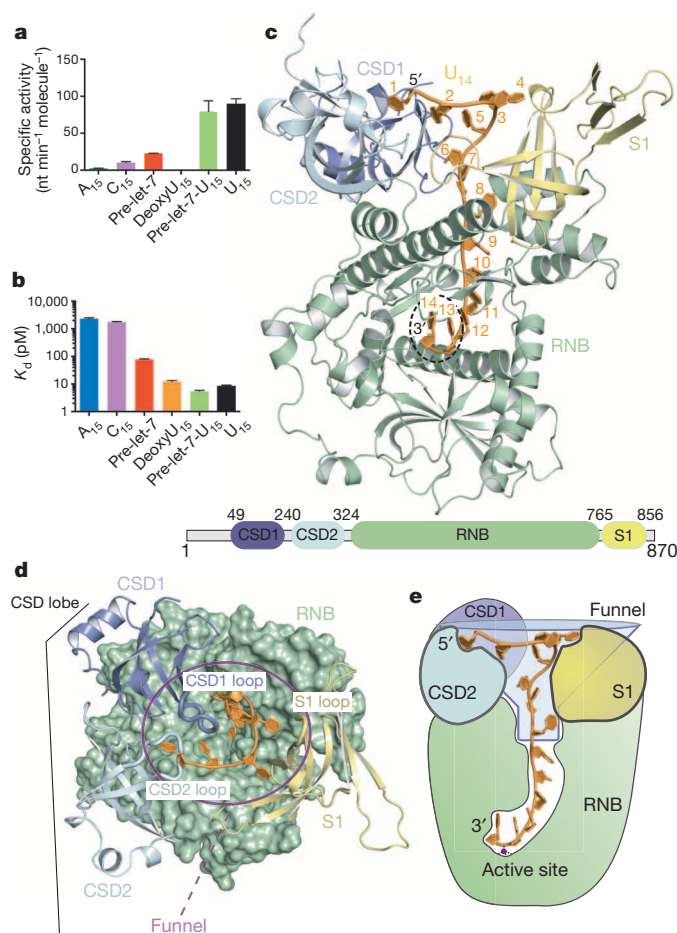


Figure 1 | Structure of Dis3l2 in complex with U₁₄ RNA. **a**, Specific activity (nucleotides (nt) min⁻¹ molecule⁻¹) of Dis3l2 with different RNA substrates. Specific activity was determined from the initial rate (fmol substrate degraded per minute) divided by the amount of enzyme (fmol) multiplied by ($n - 3$) (where n is the number of nucleotides of the substrate and 3 is the length of the end product). Mean \pm s.d. ($n = 3$) are shown. **b**, The equilibrium dissociation constant, K_d , determined by slot blot filter binding assay with different substrates conducted in triplicate. The $K_d \pm$ s.d. for each substrate is plotted. **c**, Overall structure of Dis3l2 in complex with U₁₄ and a domain schematic. Every domain of Dis3l2 makes contact to the RNA. CSD1, CSD2 (CSD lobe) and S1 are oriented in an open arrangement to form a funnel, where the 5' nucleotides U1–U8 bind, funnelling U9–U14 within the core of the catalytic RNB domain. U13 and U14 are positioned in the active site at the bottom of the RNB (black circle). **d**, Overall view of the 'top' of the Dis3l2–U₁₄ structure, rotated 90° around the horizontal axis compared to panel c. The RNB is shown in surface view. CSD1 and CSD2 form a lobe (CSD lobe) opposite S1 and are rendered as cartoon representations. All domains are colour-coded as in panel c. A purple circle represents the mouth of the CSD–S1 funnel. CSD1, CSD2 and S1 make contact to U₁₄ with residues primarily located on loops that protrude into the mouth of the funnel. **e**, Cartoon model of the funnel-shaped Dis3l2–U₁₄ complex.

Dis3l2 is structurally similar to yeast Rrp44 (refs 23–25) except for its lack of an N-terminal PIN domain, which mediates yeast Rrp44 association with the exosome (Fig. 2a, b). Both Dis3l2 and yeast Rrp44 resemble *Escherichia coli* RNase II²⁶ (Extended Data Fig. 6), and share processive 3' to 5' exonuclease activity, catalysed by a universally conserved Mg²⁺-dependent active site (Extended Data Fig. 6a, b). Individually, the CSD1 (42% sequence identity, 0.9 Å root mean squared deviation (r.m.s.d.)), CSD2 (18% sequence identity, 1.5 Å r.m.s.d.) and S1 (27% sequence identity, 1.9 Å r.m.s.d.) domains of Dis3l2 align well with the corresponding domains in yeast Rrp44 and are positioned similarly on one face of the RNB domain (39% sequence identity, 1.4 Å r.m.s.d.) (Fig. 2a, b). However, yeast Rrp44 has a narrower pore on the top of its

RNB compared to Dis3l2, due to closure of its CSD lobe over the top of the RNB to the S1. Consequently, RNA enters Rrp44 through a 'side' path formed between CSD1 and the RNB^{23,24} as it exits the exosome core (Exo10) (Fig. 2b). For Dis3l2, the funnel formed by the CSD lobe and S1 closes the side path taken in Rrp44, making it inaccessible to RNA substrates. Furthermore, the mouth of the funnel is lined with a positively charged electrostatic surface that provides an appropriate binding site for RNA substrates (Fig. 2c, d). The path taken by U₁₄ in Dis3l2 is more reminiscent of the RNA path in the structure of RNase II²² (Extended Data Fig. 6c–f). However, the opening at the top between the CSD lobe and S1 domain is narrower in the single-strand-specific RNase II enzyme²⁶ (Extended Data Fig. 6d, f), which precludes structured RNA substrates from entering.

The substrate preference exhibited by Dis3l2 is explained by numerous specific interactions with the uracil base of an oligoU-tailed RNA substrate (Fig. 3a, b). While the path that the RNA must take to the RNB domain is different in Dis3l2 compared to yeast Rrp44, the paths converge from U9 to U14 within the narrow pore of the RNB domain (Fig. 2a, b). In this region, Dis3l2 residues primarily contact the RNA backbone and are generally conserved with yeast Rrp44 (Fig. 3b). However, where the RNA paths diverge (U1–U8), we identified an extensive network of interactions between Dis3l2 and the uracil bases of U₁₄ that are not observed in structures of yeast Rrp44 (refs 23, 24) and RNase II²⁶ (Fig. 3b). We categorized these interactions into three U-specificity zones (U-zone 1–3) that together discriminate uracil from adenine and cytosine (Fig. 3c–d and Extended Data Fig. 7). U-zone 1, comprised of residues from the CSD lobe and S1 at the mouth of the funnel, includes U1–U4 (Fig. 3b, c). U1 is stacked between the side chains of R275 and F80, and forms hydrogen bonds with the side chain of H271 and the main chain of H271 and R275. U2 is stacked with H271 and is held in place by main-chain hydrogen bonds with F80 and P77. U3 and U4 interact exclusively with residues from the S1 domain (Fig. 3b, c). U3 is stacked on Y794 and interacts with the side chain of N796. U4 forms hydrogen bonds with the side chain of N777 and main chain of A779 and Q778 (Fig. 3b, c).

The bases of U6–U8 are located in the narrow stem of the funnel at the interface between the CSD lobe, S1 and RNB to make up U-zone 2 (Fig. 3b, d). U6 and U7 mediate the interaction between CSD1 and the RNB (Extended Data Fig. 4a). They are stacked and sandwiched by F84 of CSD1 and M615 of the RNB. The CSD1 side chain of R74 interacts with U6 while the side chains of CSD1 D93 and RNB Q612 contact U7. U8 is located in a pocket at the RNB–S1 interface (Extended Data Fig. 4c), formed by Q551 and N661 of the RNB and Q790 from S1. The side chains of Q551 and Q790 make direct hydrogen bonds with U8, and are stabilized by the side chain of N661 (Fig. 3b, d).

Finally, U-zone 3 includes residues from the RNB core that interact with U9–U12 (Fig. 3b, e). The base edge of U9 is engaged in Watson–Crick-like pairing with the main chain of L549 and Q551. An extended network of interactions is made between U10–U12 with the main chain of L549 and side chains of D550 and K553 (Fig. 3b, e).

Altogether we identified 22 U-specific hydrogen bonds (9 from main-chain atoms and 13 from side-chain interactions) between Dis3l2 and the uracil bases of the first 12 nucleotides of U₁₄. Most of these interactions are disrupted when we modelled A₁₄ and C₁₄ into the Dis3l2 structure (Extended Data Fig. 7), effectively providing an explanation for why oligoA and oligoC are poor substrates compared to oligoU. To examine the role of Dis3l2 residues in each U-zone, we mutated selected residues and measured activity with U₁₅ (Fig. 3f and Extended Data Fig. 8). U-zone 1 residues primarily interact with U1–U2 through main-chain atoms, which explains why side-chain mutations displayed moderate impairment (R275A). Mutations of the side chains in the S1 domain that interact with U3–U4 were more variable. N796A displayed wild-type levels of activity, whereas mutation of the same residue to its human DIS3 counterpart (N796E) is impaired (Fig. 3c, f). We suspect that the longer, negatively charged residue at this position impedes RNA progression through the enzyme compared to alanine. Mutation of Q778A

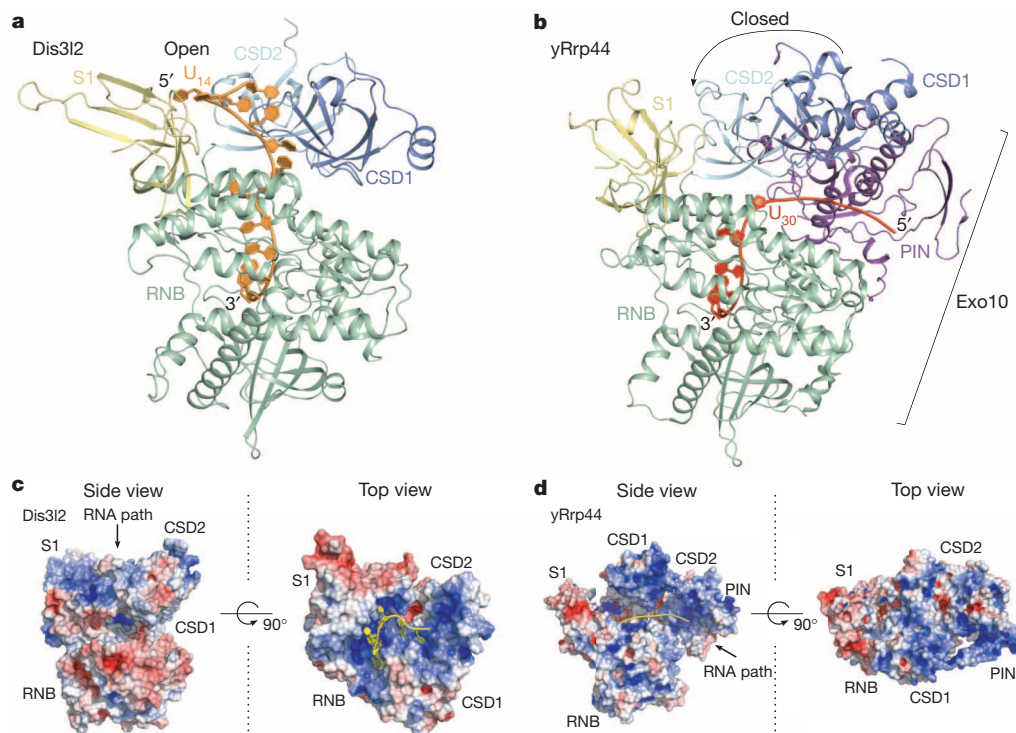


Figure 2 | Comparison of Dis3l2 to yeast Rrp44 of the RNA exosome.

a, Dis3l2–U₁₄ complex rotated 180° around the vertical axis compared to Fig. 1c. The open funnel created by the CSD lobe and S1 allows RNA to access the top of the RNB. **b**, The structure of yeast Rrp44 extracted from the 11-subunit RNA exosome²⁰ (Protein Data Bank 4IFD). The CSD lobe closes on the S1 domain leading to a closed conformation on top of the RNB; as a result a ‘side’ RNA binding path is created between the RNB and CSD1.

c, Perpendicular side and top views of the electrostatic surface potential of Dis3l2 (contoured at $\pm 5 \text{ kT } e^{-1}$; white indicates neutral, blue indicates positive and red indicates negative). A positively charged electrostatic surface lines the wide portion of the funnel on the top of the RNB that can accommodate structured RNA substrates. **d**, Electrostatic surface potential of the exosome-associated yeast Rrp44, in the same configuration as panel c, supports an RNA-binding path along the side route.

is the most defective in U-zone 1 and we speculate that Q778 might be involved in the translocation of U3 to the U4 position (Fig. 3 c, f). Unexpectedly, we identified an activating mutation (N777A) in U-zone 1 that accounts for an enzyme with 40% higher levels of activity than wild-type Dis3l2. U4 is held in place through main-chain interactions in addition to the side chain of N777, so that mutation to a small aliphatic residue maintains part of the interaction, but may create a hollow path to allow RNA to pass unimpeded to U-zone 2 (Fig. 3c, f). Similar ‘super-enzyme’ mutations of functional to small aliphatic residues have also been described for RNase II²⁷ and yeast Rrp44 (ref. 28).

Notably, two of the most impaired mutants (R74A/Q612A and Q790A) in U-zone 2 are of side chains that not only read the RNA sequence but also engage in oligoU-mediated domain–domain interfaces (Fig. 3d, f and Extended Data Fig. 4a). R74 and Q612 stabilize the stacked conformation of U6–U7 and sit at the junction of the RNA stabilized CSD1–RNB interface. Q790 facilitates an RNA-mediated interface between the S1 and RNB domains (Fig. 3d, f). Collectively, these data strongly support the U₁₄ path and the domain configuration in the Dis3l2 structure.

U-zone 3 seems to be the least U-specific zone, as we had to mutate a stretch of six residues in Dis3l2 (R548–K553) to the mammalian Dis3 sequence (Fig. 3e, f and Extended Data Fig. 2) to achieve a 40% reduction in activity. U-zone 3 reads uracil through a mixture of main chain and side chains, but is mostly composed of non-sequence-specific interactions with the U₁₄ backbone (Fig. 3b), which may explain the relatively modest effect on activity for this mutant.

All of the U-zone mutants examined are active, which is not entirely surprising given the extensive network of interactions maintained throughout all of the U-zones, even in the context of a single point mutation. This is particularly evident when we compared the specific activity of any single U-zone mutation, where 1–2 interactions are broken, with

A₁₅ or C₁₅, where the bulk of the U-zone network is disrupted (Fig. 3f and Extended Data Fig. 7). To clarify the impact of U-zone mutations on Dis3l2’s processing mechanism, we measured the binding affinity for U₁₅ (Fig. 3g and Extended Data Fig. 9). There is little correlation between binding and activity here, given that U-zone mutations had little impact on the binding affinity for U₁₅ compared to wild-type Dis3l2 (Fig. 3g). Because single U-zone mutations are not sufficient to abolish oligoU binding, again because of an overwhelming network of interactions, it may rather be the rate of oligoU translocation through the mutated U-zones to the active site that is impeded.

The structure of Dis3l2 presented here answers several important mechanistic questions about the biochemical function of this essential enzyme. Most importantly, our study identified a vast network of oligoU-specific interactions, even when compared to a typical transcription factor binding to its recognition sequence. This accounts for Dis3l2’s enhanced activity and higher affinity for oligoU-tailed substrates (Fig. 1a, b). The shape of the binding funnel—wide at the top and narrow at the bottom (Fig. 1c–e and Fig. 2a, c)—explains its ability to process structured RNA substrates. Indeed, Dis3l2 is more adept at processing structured RNA, even those with short 3′-end overhangs, compared to human DIS3 (ref. 20) and yeast Rrp44 (ref. 20). We propose that the CSD–S1 funnel is wide enough to bind structured RNA. In turn, the CSDs and S1, which are known to function as RNA chaperones in other proteins²⁹, may promote the remodelling of structured RNA. Our structure supports this model quite well, as we observe a far-reaching set of base interactions, including base-stacking and hydrophilic interactions, throughout the CSD–S1 funnel that could have this role (Fig. 3b–e). Once the U-zones are primed with an oligoU tail, Dis3l2 will degrade it up through the RNB, where non-sequence-specific interactions (Fig. 3b) and hydrolysis in the active site fuel translocation of RNA substrates, like pre-let-7,

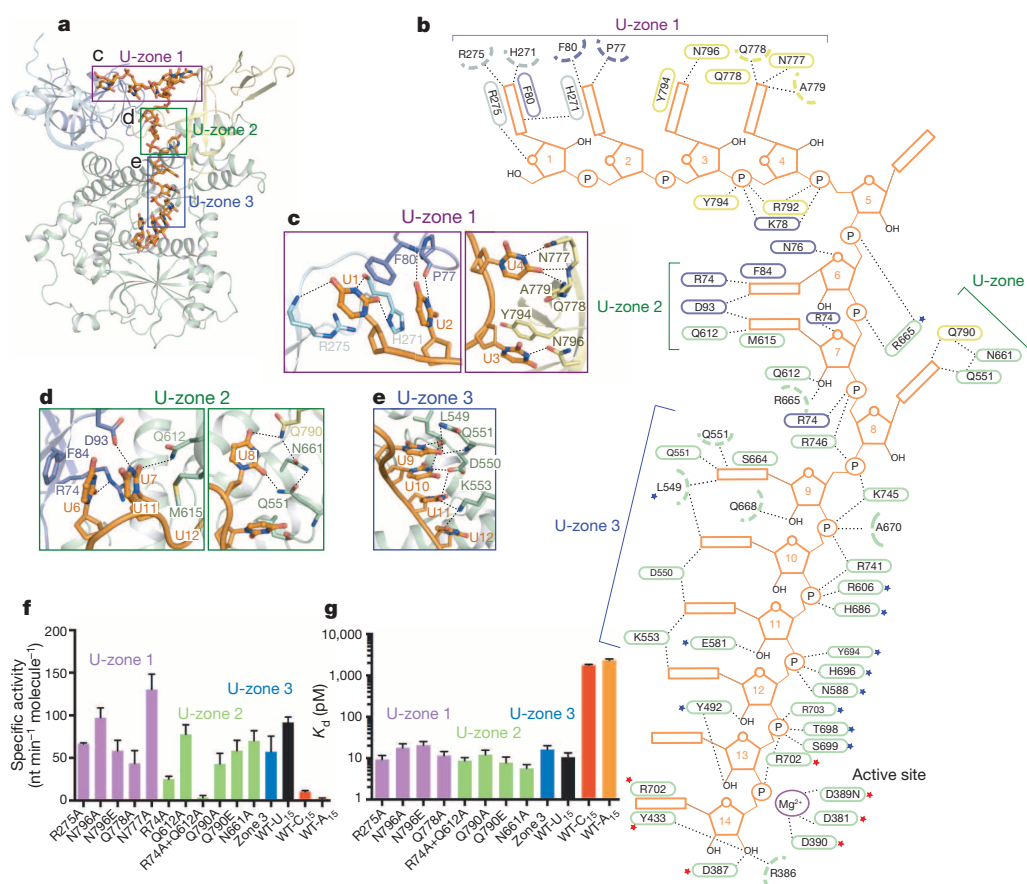


Figure 3 | Mechanism of Dis3L2's oligoU specificity. **a**, Overall structure of Dis3L2–U₁₄ with the location of three uracil specificity zones (U-zone 1–3) labelled and coloured purple (U-zone 1), green (U-zone 2) and blue (U-zone 3). **b**, A schematic view of the interactions between U₁₄ and Dis3L2. The interactions between U₁₄ (orange) with side chain (ovals) and main chain (dashed semi-circles) residues of Dis3L2 are labelled and colour-coded by domain. Dashed lines represent hydrogen bonds and/or ion pairs. The active site of Dis3L2 is labelled. Active site residues conserved with yeast Rrp44 are highlighted with red asterisks. Conserved residues that interact with the RNA backbone are labelled with blue asterisks. U-zones are labelled as in panel **a**. **c–e**, Detailed U-specific interactions in U-zones 1–3 between Dis3L2 and the

uracil bases of U1–U12. **c**, The bases of U1–U4 in U-zone 1. The uracil bases are shown as orange sticks and Dis3L2 residues are sticks, colour-coded by domain as in Fig. 1c. Hydrogen bonds are shown as dashed lines. **d**, U-zone 2 encompassing U6–U8. **e**, U9–U12 in U-zone 3. **f**, Specific activity (nt min^{−1} molecule^{−1}) of selected U-zone mutants with U₁₅ substrate compared to wild type with U₁₅, C₁₅ and A₁₅. Specific activity was measured with 0.5 nM of indicated mutant and 100 nM U₁₅. Mean ± s.d. (*n* = 3) are shown. WT, wild type. **g**, Equilibrium dissociation constants, *K_d*, of U-zone mutants with U₁₅, compared to wild type with U₁₅, C₁₅ and A₁₅. All assays were conducted in triplicate. The *K_d* ± s.d. for each mutant is plotted.

through the enzyme. In conclusion, our data provide the structural mechanism of substrate recognition that underlies Dis3L2's role as the effector in maintaining pluripotency via the Lin28–let-7 pathway.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 April; accepted 2 June 2014.

Published online 3 August 2014.

- Viswanathan, S. R., Daley, G. Q. & Gregory, R. I. Selective blockade of microRNA processing by Lin28. *Science* **320**, 97–100 (2008).
- Heo, I. *et al.* Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. *Mol. Cell* **32**, 276–284 (2008).
- Rybak, A. *et al.* A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment. *Nature Cell Biol.* **10**, 987–993 (2008).
- Newman, M. A., Thomson, J. M. & Hammond, S. M. Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. *RNA* **14**, 1539–1549 (2008).
- Thornton, J. E. & Gregory, R. I. How does Lin28 let-7 control development and disease? *Trends Cell Biol.* **22**, 474–482 (2012).
- Zhu, H. *et al.* The Lin28/let-7 axis regulates glucose metabolism. *Cell* **147**, 81–94 (2011).
- Shyh-Chang, N. *et al.* Lin28 enhances tissue repair by reprogramming cellular metabolism. *Cell* **155**, 778–792 (2013).
- Urbach, A. *et al.* Lin28 sustains early renal progenitors and induces Wilms tumor. *Genes Dev.* **28**, 971–982 (2014).

- Nam, Y., Chen, C., Gregory, R. I., Chou, J. J. & Sliz, P. Molecular basis for interaction of let-7 microRNAs with Lin28. *Cell* **147**, 1080–1091 (2011).
- Hagan, J. P., Piskounova, E. & Gregory, R. I. Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells. *Nature Struct. Mol. Biol.* **16**, 1021–1025 (2009).
- Heo, I. *et al.* TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. *Cell* **138**, 696–708 (2009).
- Thornton, J. E., Chang, H. M., Piskounova, E. & Gregory, R. I. Lin28-mediated control of let-7 microRNA expression by alternative TUTases Zcchc11 (TUT4) and Zcchc6 (TUT7). *RNA* **18**, 1875–1885 (2012).
- Chang, H. M., Triboulet, R., Thornton, J. E. & Gregory, R. I. A role for the Perlman syndrome exonuclease Dis3L2 in the Lin28–let-7 pathway. *Nature* **497**, 244–248 (2013).
- Ustianenko, D. *et al.* Mammalian DIS3L2 exonuclease targets the uridylated precursors of let-7 miRNAs. *RNA* **19**, 1632–1638 (2013).
- Liu, Q., Greimann, J. C. & Lima, C. D. Reconstitution, activities, and structure of the eukaryotic RNA exosome. *Cell* **127**, 1223–1237 (2006).
- Dziembowski, A., Lorentzen, E., Conti, E. & Seraphin, B. A single subunit, Dis3, is essentially responsible for yeast exosome core activity. *Nature Struct. Mol. Biol.* **14**, 15–22 (2007).
- Tomecki, R. *et al.* The human core exosome interacts with differentially localized processive RNases: hDIS3 and hDIS3L. *EMBO J.* **29**, 2342–2357 (2010).
- Astuti, D. *et al.* Germline mutations in DIS3L2 cause the Perlman syndrome of overgrowth and Wilms tumor susceptibility. *Nature Genet.* **44**, 277–284 (2012).
- Malecki, M. *et al.* The exonuclease Dis3L2 defines a novel eukaryotic RNA degradation pathway. *EMBO J.* **32**, 1842–1854 (2013).
- Lubas, M. *et al.* Exonuclease hDIS3L2 specifies an exosome-independent 3'–5' degradation pathway of human cytoplasmic mRNA. *EMBO J.* **32**, 1855–1868 (2013).
- Chang, H., Lim, J., Ha, M. & Kim, V. N. TAIL-seq: Genome-wide determination of poly(A) tail length and 3' end modifications. *Mol. Cell* **53**, 1044–1052 (2014).

22. Choi, Y. S., Patena, W., Leavitt, A. D. & McManus, M. T. Widespread RNA 3'-end oligouridylation in mammals. *RNA* **18**, 394–401 (2012).
23. Lorentzen, E., Basquin, J., Tomecki, R., Dziembowski, A. & Conti, E. Structure of the active subunit of the yeast exosome core, Rrp44: diverse modes of substrate recruitment in the RNase II nuclease family. *Mol. Cell* **29**, 717–728 (2008).
24. Makino, D. L., Baumgartner, M. & Conti, E. Crystal structure of an RNA-bound 11-subunit eukaryotic exosome complex. *Nature* **495**, 70–75 (2013).
25. Bonneau, F., Basquin, J., Ebert, J., Lorentzen, E. & Conti, E. The yeast exosome functions as a macromolecular cage to channel RNA substrates for degradation. *Cell* **139**, 547–559 (2009).
26. Frazão, C. *et al.* Unravelling the dynamics of RNA degradation by ribonuclease II and its RNA-bound complex. *Nature* **443**, 110–114 (2006).
27. Barbas, A. *et al.* Determination of key residues for catalysis and RNA cleavage specificity: one mutation turns RNase II into a “SUPER-ENZYME”. *J. Biol. Chem.* **284**, 20486–20498 (2009).
28. Reis, F. P. *et al.* Modulating the RNA processing and decay by the exosome: altering Rrp44/Dis3 activity and end-product. *PLoS ONE* **8**, e76504 (2013).
29. Rajkowitsch, L. *et al.* RNA chaperones, RNA annealers and RNA helicases. *RNA Biol.* **4**, 118–130 (2007).

Acknowledgements We thank C. M. Hammell for comments on this manuscript and members of the Joshua-Tor laboratory for comments and suggestions. We thank A. Héroux for help at the National Synchrotron Light Source, which is supported by Department of Energy, Office of Basic Energy Sciences. We also thank the Protein Core Facility at Columbia University. This work was supported by the Watson School of Biological Sciences (to J.W. and L.J.), the Louis Morin Charitable Trust and the Robertson Research Fund of Cold Spring Harbor Laboratory (to L.J.). L.J. is an investigator of the Howard Hughes Medical Institute.

Author Contributions C.R.F., J.W. and L.J. designed and C.R.F. and J.W. conducted all experiments. All authors contributed to data analysis and wrote the paper.

Author Information Coordinates and structure factors for the Dis3l2–U₁₄ complex have been deposited in the Protein Data Bank under accession code 4PMW. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.J. (leemor@cshl.edu).

METHODS

Protein preparation. Mouse Dis3l2 was expressed in Sf9 cells as an N-terminal Strep-sumo-TEV fusion protein from the pFL vector of the MultiBac baculovirus expression system³⁰. Sf9 cells were infected with baculovirus in Hyclone CCM3 media (Thermo Scientific) at 27 °C. Following 60 h of expression, the cells were centrifuged at 1,200 r.p.m. and re-suspended in Wash buffer (50 mM Tris pH 8, 100 mM NaCl and 5 mM DTT), flash frozen in liquid nitrogen and stored at −80 °C. Frozen cells were thawed, NaCl concentration increased to 500 mM, and then lysed by sonication. The lysate was treated with 0.2% poly-ethylene imine (PEI) to precipitate nucleic acids before ultracentrifugation at 35,000 r.p.m. at 4 °C for 1 h. The soluble fraction was incubated with 1 ml of Strep-Tactin superflow resin (IBA bioTAGnology) per 10 ml of lysate for 1 h on a rolling shaker. The resin was applied to a gravity flow column and washed extensively with Wash buffer. The protein was eluted with Wash buffer containing 2 mM desthiobiotin. The eluted fraction was treated with TEV protease overnight at 4 °C. The cleavage efficiency and purity was verified by SDS-PAGE. The cleaved protein was diluted with an equal volume of heparin buffer A (25 mM HEPES pH 7.5, 5 mM DTT) to a final NaCl concentration of 50 mM. Dis3l2 was loaded onto a HiTrap heparin HP column (GE Life Science) equilibrated with 25 mM HEPES pH 7.5, 50 mM NaCl and 5 mM DTT. A linear gradient between 0.05 M and 1 M NaCl was used to elute Dis3l2 at 0.25 M NaCl. Fractions that contained Dis3l2 were analysed by SDS-PAGE, pooled and concentrated to 2 ml and loaded onto a HiLoad 16/60 Superdex 200 gel filtration column equilibrated with 10 mM Tris pH 8, 100 mM NaCl, 2 mM MgCl₂ and 5 mM DTT. Fractions containing Dis3l2 were pooled and concentrated to 25 mg ml^{−1}, flash frozen in liquid nitrogen and stored at −80 °C. Mutants of Dis3l2 were constructed by sequence and ligation-independent cloning³¹ (SLIC) using mutant primers. All mutant proteins were expressed and purified as described for wild-type Dis3l2.

Crystallization. Full-length Dis3l2 was recalcitrant to crystallization despite extensive screening. We identified a protease-sensitive loop within CSD1 by limited proteolysis with thermolysin. We expressed and purified a D389N active-site mutant lacking the protease-sensitive loop within CSD1 (residues 148–169 and 194–221), and both the N (1–36) and C (857–870) termini (Extended Data Fig. 2). The truncated Dis3l2 was incubated with a 1.2 molar excess of U₁₃ RNA for 30 min at 20 °C. Crystallization was carried out by the hanging-drop vapour diffusion method by mixing the Dis3l2–U₁₃ complex at 15 mg ml^{−1} with an equal volume of 100 mM tri-ammonium citrate pH 5.5, 100–200 mM ammonium chloride and 20–22% PEG 3350. Crystals appeared in 1–2 days at 18 °C and were placed directly in a freshly prepared solution composed of the crystallization condition supplemented with 20% glycerol and flash frozen in liquid N₂.

Structure determination. X-ray diffraction data were collected to 2.95 Å resolution at the X25 beamline at the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory (BNL) (Extended Data Table 1). Data were processed with XDS³² as implemented in the autoPROC software³³. We generated a molecular replacement (MR) search model by pruning yeast Rrp44 (ref. 23) (PDB 2VNU) with the Sculptor³⁴ utility. The pruned model was separated into its individual CSD1, CSD2, RNB and S1 domains. Phasing was determined by MR with Phaser³⁵, using each of the domains of yeast Rrp44 as individual search models. Phaser found solutions for CSD2, RNB and S1 domains, but failed to find a unique solution for CSD1. The initial MR phases were input into the AutoBuild routine in Phenix³⁶, where the Resolve³⁷ and Buccaneer³⁸ option for density modification and model building, respectively, was critical for phase improvement and interpretation of initial electron density maps. The CSD1 of yeast Rrp44 was placed manually into the initial map and the AutoBuild routine repeated. The resulting phases were of sufficient quality to manually build protein and RNA in Coot³⁹. Although a U₁₃ homooligonucleotide was used in crystallization, interpretable electron density accounted for 14 nucleotides. This is readily explained by slipping of the homo-oligoU into two conformations along the RNA binding cleft. One conformation is fully threaded into the active site (at the 0 position), and a second conformation occupies the −1 position. Crystallographically, these two states are indistinguishable, and thus is best accounted for by modelling 14 nucleotides. The final model includes two copies of truncated Dis3l2 with the exception of disordered residues (37–48, 119–228 and 252–258), two U₁₄ molecules, two Mg²⁺ ions and six water molecules. The structure was refined using automatically determined NCS restraints, TLS and isotropic B-factor refinement as implemented in Phenix⁴⁰ to an R/R_{free} of 21/25%. The final model was validated with MolProbity⁴¹, where 96.8% of residues reside in the favoured and 3.13% in the allowed regions of the Ramachandran plot (Extended Data Table 1). The copy represented by chain A (Dis3l2) and chain C (U₁₄) had appreciably better electron density and was used for all structural analysis described in this manuscript. Weak electron density, presumably due to disorder in the region surrounding U5, precluded accurate modelling of the position of this nucleotide,

so we carefully avoided over interpretation of the interactions observed in this region, although we are confident of its presence and general location. All structural figures were generated with the PyMOL Molecular Graphics System, Version 1.6 Schrödinger, LLC.

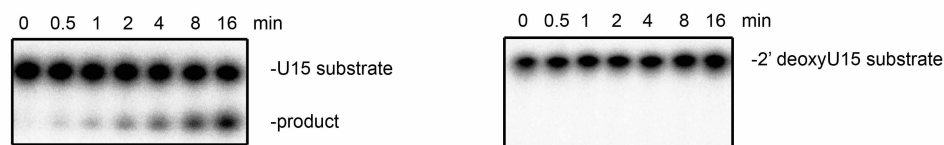
Specific activity determination. All exonuclease assays were conducted under multiple turnover conditions, where we measured the linear initial rate of RNA degradation and calculated the specific activity (nt degraded min^{−1} molecule of enzyme^{−1}). The indicated amount of Dis3l2 was incubated at 30 °C with 100 nM 5' radiolabelled RNA substrate in reaction buffer (20 mM HEPES pH 7.5, 50 mM NaCl, 1 mM DTT and 100 μM MgCl₂). A 50 μl reaction was initiated and 5 μl was removed and quenched at 0, 0.5, 1, 2, 4, 8 and 16 min time points in formamide loading buffer (0.025% SDS, 95% formamide, 0.025% bromophenol blue, 0.025% xylene cyanol and 18 mM EDTA). Quenched samples were heated to 95 °C for 2 min and resolved by denaturing urea PAGE. Products were analysed by phosphor imaging and quantified with ImageJ software. The amount of substrate degraded (fmol) was plotted against time and linear regression was used to determine the initial rate (slope) of the linear portion of the curve with GraphPad Prism 6. The initial rate was converted to specific activity (nt/min/molecule of enzyme) by dividing the initial rate (fmol substrate processed/min) by the fmol of enzyme used in each assay and multiplying by $n - 3$ (where n is the number of nucleotides of the substrate and 3 is the length of the end product). Synthetic oligoU, oligoA and oligoC substrates were purchased from Trilink BioTechnologies. Pre-let-7 and pre-let-7-U₁₅ were *in vitro* transcribed. Briefly, the RNA coding sequence was flanked by a 5' hammerhead ribozyme and 3' hepatitis delta virus ribozyme (HDV) to ensure homogeneous ends^{42,43}. Constructs were cloned into a pRSF vector containing a T7 promoter and the RNA produced by run-off transcription. The pre-let-7 sequence used in this study is pre-let7a1: UGAGGUAGUAGGUUGUAUAGUUUAGGGUCACAC CCACCACUGGGAGAUAAUAUACAAUCUACUGUCUUUC.

Transcribed RNA was gel purified with denaturing PAGE and re-suspended in DEPC-treated water.

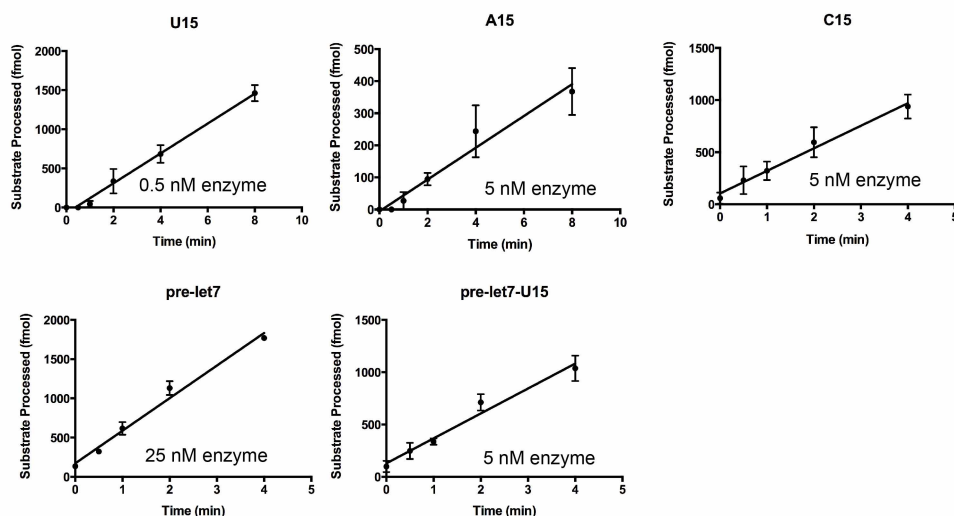
Equilibrium binding assays. A range of Dis3l2 concentrations—0–5 nM for oligoU, pre-let-7-U₁₅, deoxyU₁₅ and pre-let-7 substrates and 0–150 nM for A₁₅ and C₁₅—were incubated with 1 pM 5' radiolabelled RNA for 1 h in binding buffer (20 mM HEPES pH 7.5, 50 mM NaCl, 1 mM DTT, 50 μg ml^{−1} BSA and 100 μM EDTA). We determined that wild-type Dis3l2 is inactive in the presence of EDTA and no magnesium, but binds to RNA with the same affinity as the inactive D389N mutant. A slot blot filtration system was used (BioRad) to capture Dis3l2–RNA complexes on the top nitrocellulose membrane and unbound RNA on the bottom nylon membrane. The membranes were washed with 100 μl of binding buffer, before applying 100 μl of the binding reaction, followed by 100 μl of binding buffer. The nitrocellulose and nylon membranes were dried and analysed by phosphor imaging. All binding assays were conducted in triplicate, were quantified and fraction bound plotted versus free protein concentration. K_d values were determined by using nonlinear regression analysis with GraphPad Prism6 software.

- Bieniossek, C., Richmond, T. J. & Berger, I. MultiBac: multigene baculovirus-based eukaryotic protein complex production. *Curr. Protoc. Protein Sci.* Ch. 5, Unit 5 20 (2008).
- Li, M. Z. & Elledge, S. J. Harnessing homologous recombination *in vitro* to generate recombinant DNA via SLIC. *Nature Methods* **4**, 251–256 (2007).
- Kabsch, W. Xds. *Acta Crystallogr. D* **66**, 125–132 (2010).
- Vonrhein, C. *et al.* Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. D* **67**, 293–302 (2011).
- Bunkóczi, G. & Read, R. J. Improvement of molecular-replacement models with Sculptor. *Acta Crystallogr. D* **67**, 303–312 (2011).
- McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
- Terwilliger, T. C. *et al.* Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. D* **64**, 61–69 (2008).
- Terwilliger, T. C. Maximum-likelihood density modification. *Acta Crystallogr. D* **56**, 965–972 (2000).
- Cowan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D* **62**, 1002–1011 (2006).
- Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
- Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
- Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, 375–383 (2007).
- Price, S. R., Ito, N., Oubridge, C., Avis, J. M. & Nagai, K. Crystallization of RNA-protein complexes. I. Methods for the large-scale preparation of RNA suitable for crystallographic studies. *J. Mol. Biol.* **249**, 398–408 (1995).
- Shechner, D. M. & Bartel, D. P. The structural basis of RNA-catalyzed RNA polymerization. *Nature Struct. Mol. Biol.* **18**, 1036–1042 (2011).

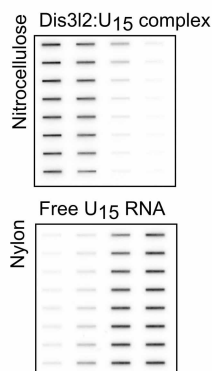
a



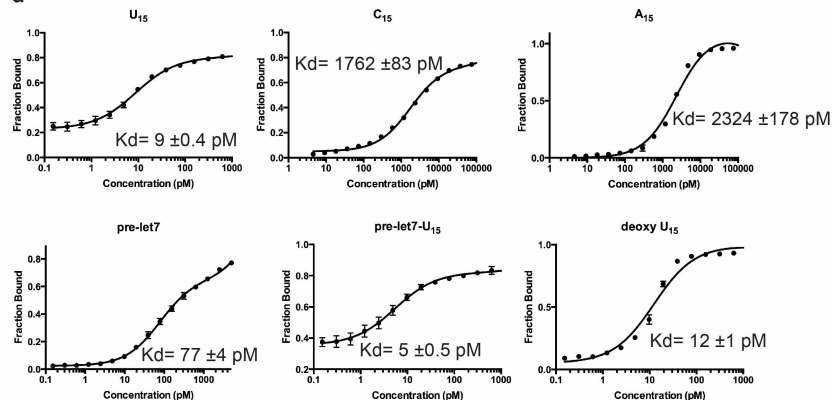
b



c

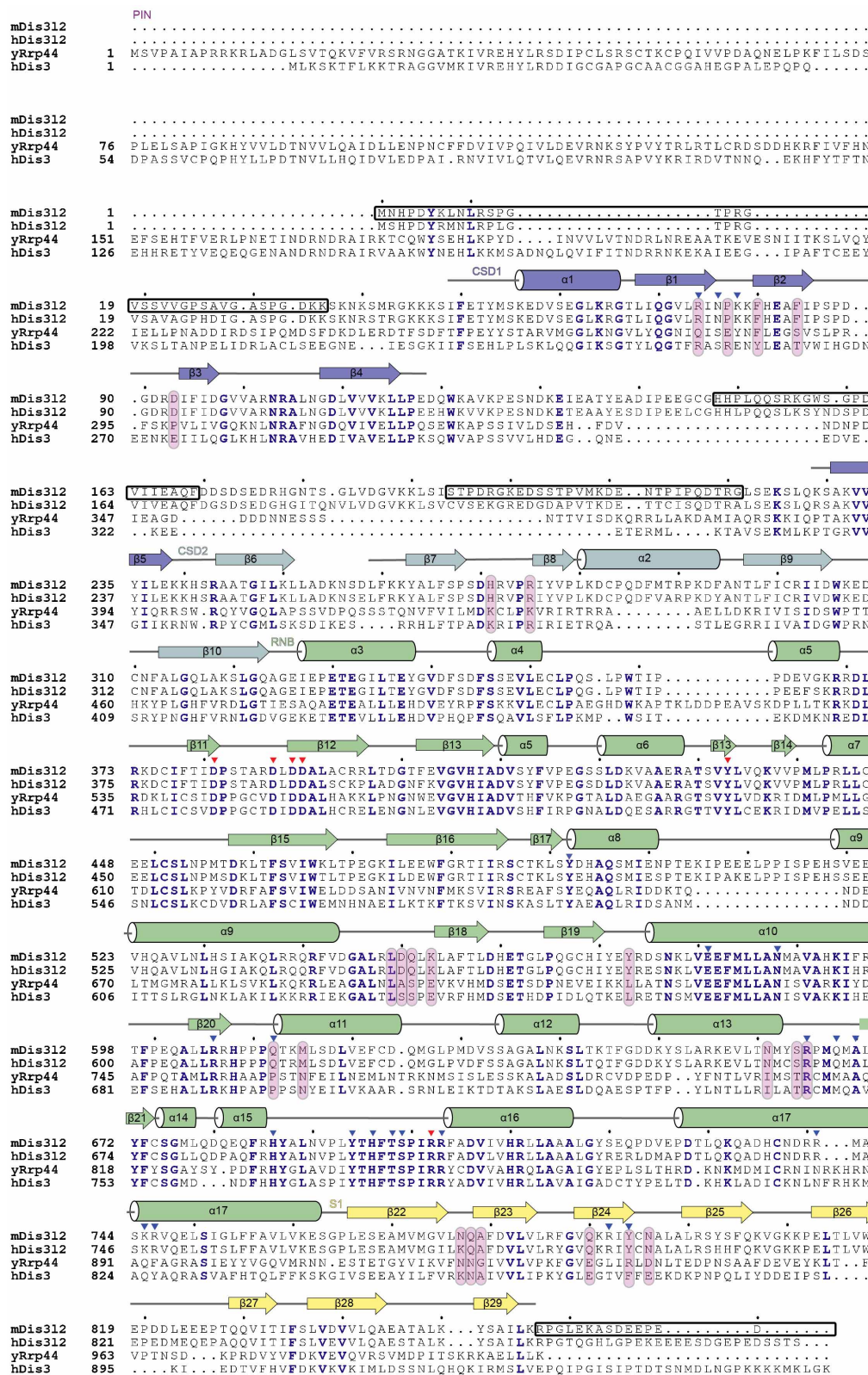


d



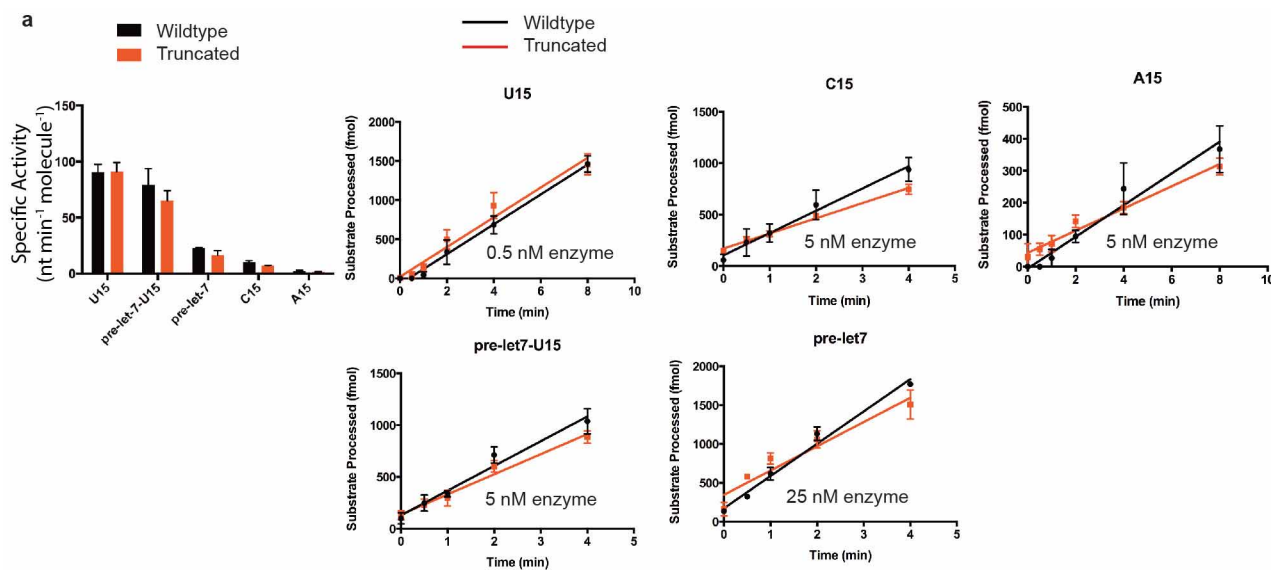
Extended Data Figure 1 | Substrate specificity of Dis3l2. **a**, Dis3l2 exonuclease assays were conducted with 5' radio-labelled RNA substrates at 30 °C over a 16 min time course and products resolved by denaturing urea PAGE. Two representative gels, one from an exonuclease assay with U₁₅ that was used to quantify Dis3l2 enzymatic activity and another for deoxyU₁₅, showing no detectable activity under the initial rate conditions tested. **b**, The initial rate plots for each substrate used to calculate specific activity as shown in Fig. 1a. Exonuclease assays were conducted in triplicate and quantified with ImageJ. The concentration of enzyme used to measure the initial rate is indicated on each plot. The amount of substrate degraded (fmol) was plotted

against time and the initial rate (slope) was determined with linear regression using GraphPad Prism 6. Mean \pm s.d. ($n = 3$) are shown. **c**, Representative slot blot filter-binding assay of Dis3l2 with U₁₅. Dis3l2–RNA complexes were captured on the top nitrocellulose membrane and unbound RNA to the bottom nylon membrane. **d**, The equilibrium dissociation constant (K_d) for each substrate in Fig. 1b was determined by plotting the fraction RNA bound against the concentration of free Dis3l2 and fit by nonlinear regression with GraphPad Prism 6. The mean of three independently measured replicates with error bars representing \pm s.d. are shown with the K_d .

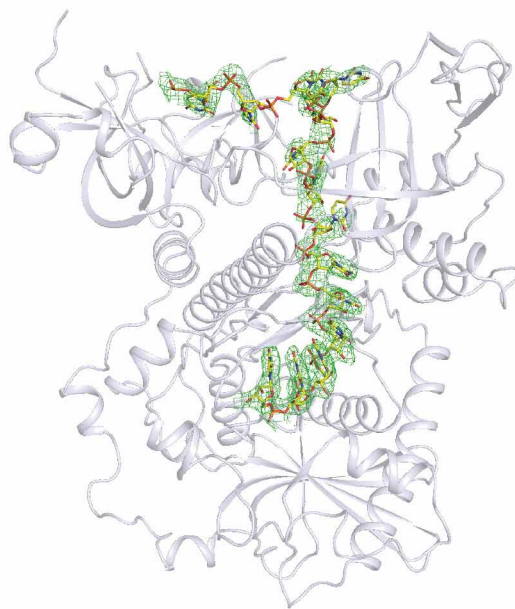


Extended Data Figure 2 | Sequence alignment of mouse and human Dis3l2, yeast Rrp44 and human Dis3. Dis3l2 is composed of two cold shock domains (CSD1, residues 49–240, $\alpha 1$, $\beta 1$ – $\beta 5$, and CSD2, residues 241–324, $\beta 6$ – $\beta 10$, $\alpha 2$) at its N terminus followed by a catalytic RNB domain (residues 325–765, $\alpha 3$ – $\alpha 17$, $\beta 11$ – $\beta 21$) and flanked by an S1 domain (residues 766–856, $\beta 22$ – $\beta 29$). The secondary structure elements deduced from the structure of mouse

Dis3l2 are shown on top of the sequences, colour-coded by domain as in Fig. 1a. The segments of mouse Dis3l2 that were truncated to facilitate crystallization are outlined in a black box. Conserved amino acid residues are coloured blue. Blue diamonds indicate residues interacting with the backbone of U₁₄ RNA and red diamonds denote conserved active-site residues. Residues in Dis3l2 that interact with U₁₄ bases are shaded purple.

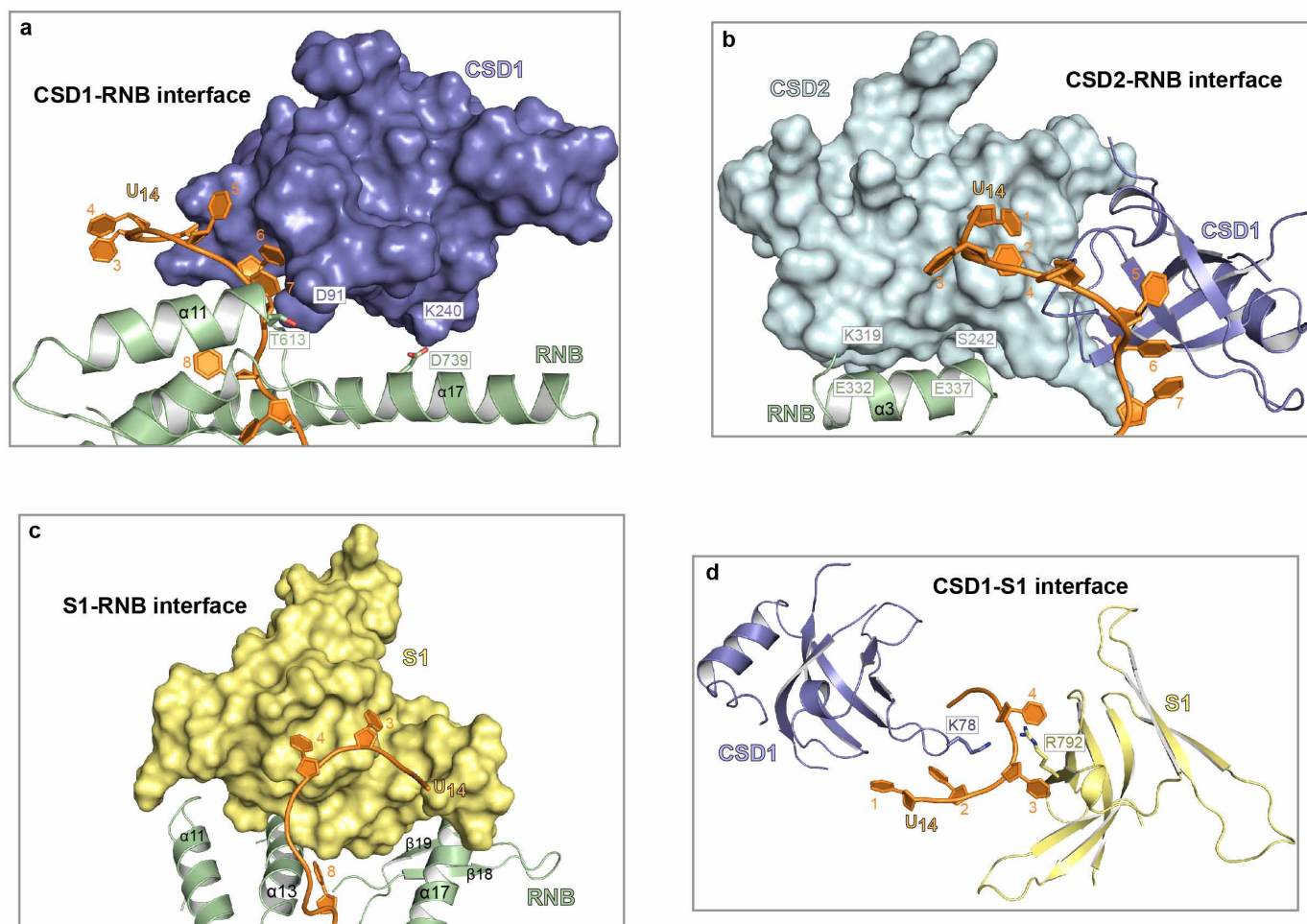


b



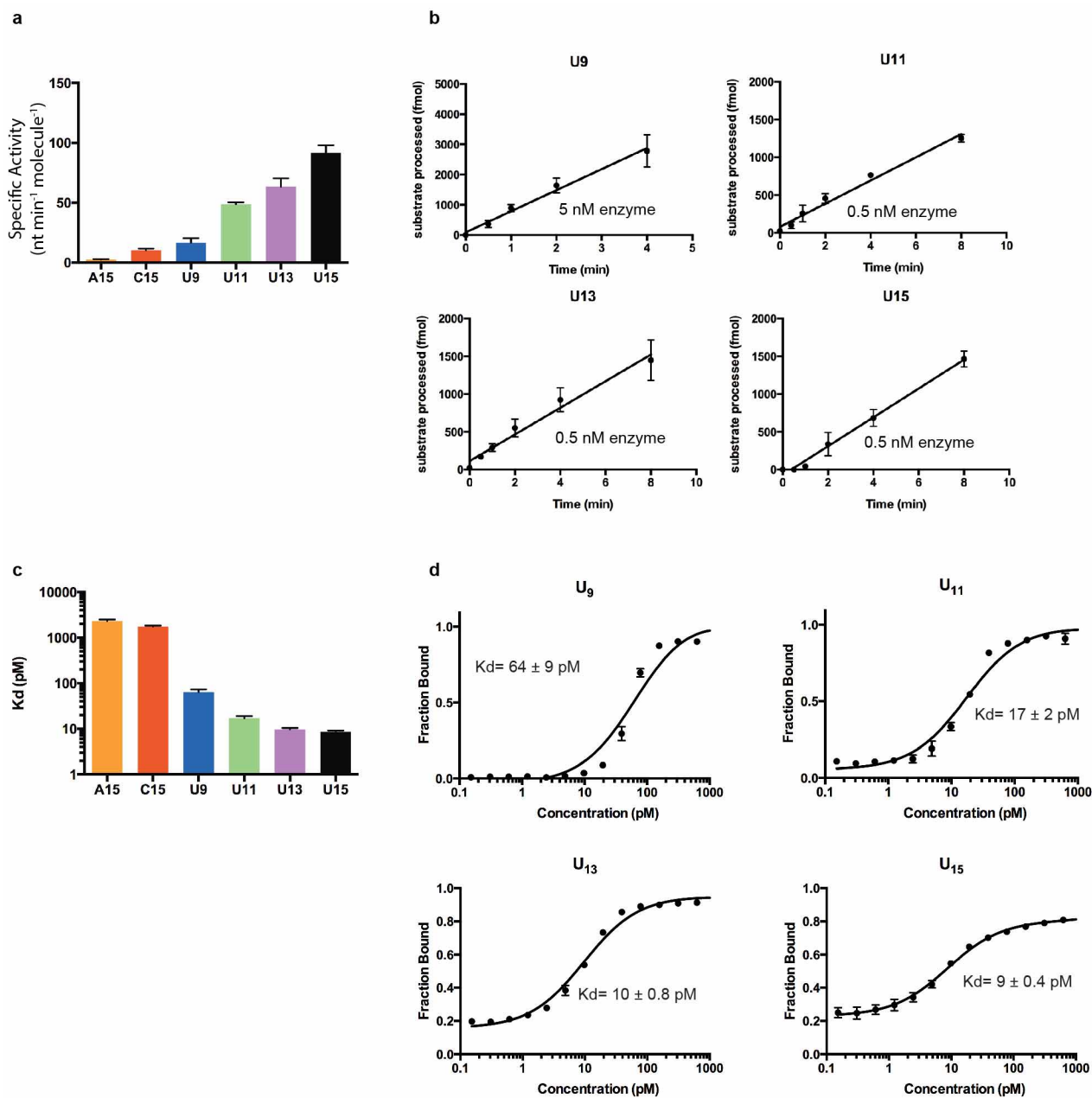
Extended Data Figure 3 | Substrate specificity of truncated Dis3l2 and electron density map for U₁₄ RNA. **a**, Comparison of specific activity ($\text{nt min}^{-1} \text{ molecule}^{-1}$) of truncated Dis3l2 construct resembling the one used in crystallization, but without the D389N mutation with the same set of substrates analysed for wild-type Dis3l2. Truncated Dis3l2 had comparable levels of activity and displayed the same substrate preference as wild-type Dis3l2. The specific activity was calculated from the initial rate plots, in the same way as described in Fig. 1a and Extended Data Fig. 1. The mean \pm s.d. ($n = 3$) are shown. **b**, The final model of Dis3l2 is shown as a transparent white

cartoon. The final model of U₁₄ is rendered as yellow sticks. Initial molecular replacement phases were improved by density modification with Resolve³² and automated model building with Buccaneer³³, as implemented by the AutoBuild wizard in Phenix³¹. The unbiased density modified electron density map before inclusion of RNA is shown contoured at 1σ . Despite the crystallization of Dis3l2 in complex with U₁₃, unambiguous density allowed modelling of U₁₄ (details are discussed in the Methods). Clear electron density accounted for all 14 nucleotides, except some disorder contributed to weak electron density surrounding U5, precluding its accurate placement in density.



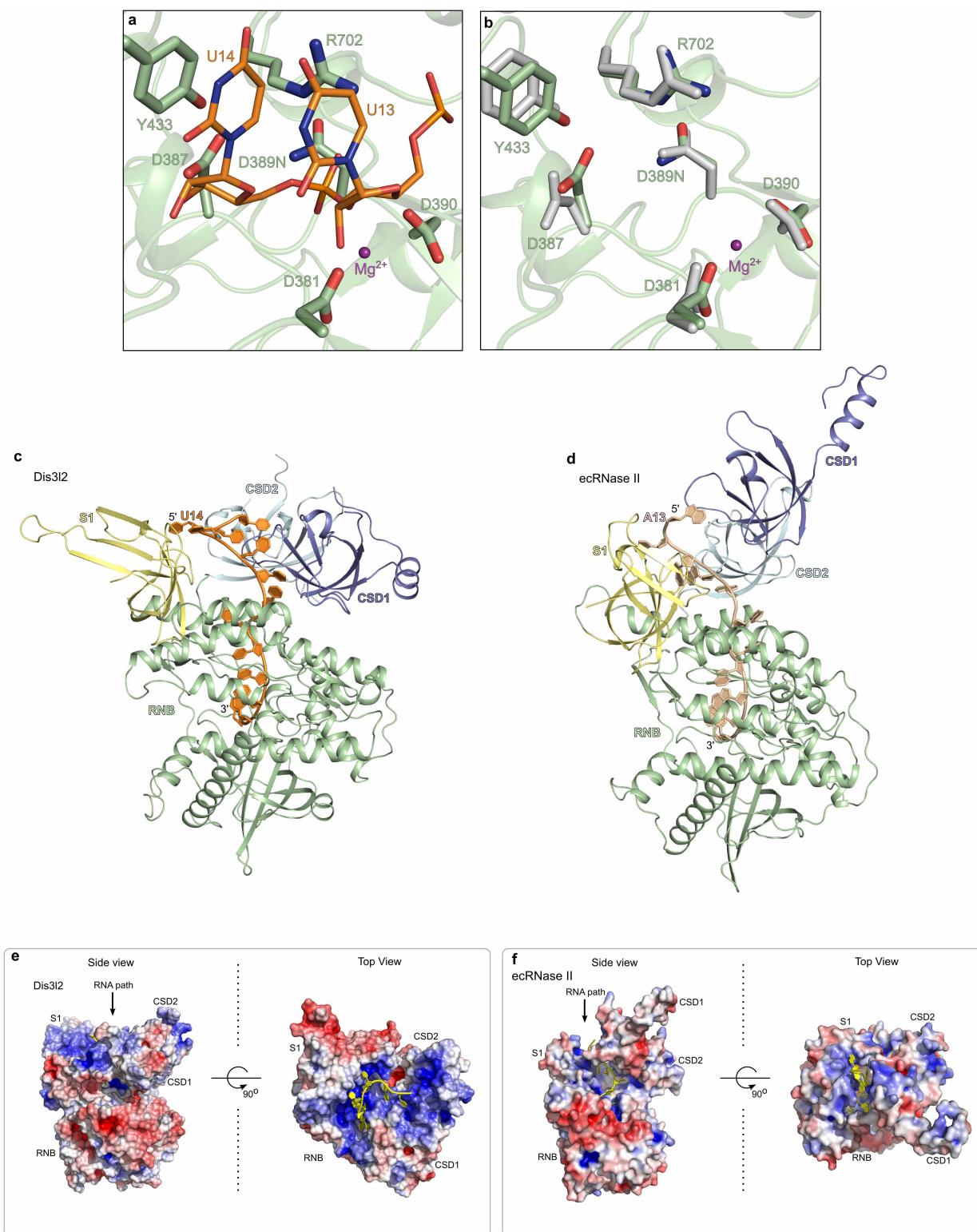
Extended Data Figure 4 | Dis3l2 domain interface analysis. **a**, Analysis of the CSD1–RNB interface. The conformation of CSD1 is stabilized by two protein–protein interactions with the RNB (K240 with D739 and D91 with T613) and an RNA-mediated interaction with the RNB through U6–U7 and $\alpha 11$. **b**, Analysis of the CSD2–RNB interface. CSD2 is intimately associated with CSD1, but also forms an interface with the RNB through $\alpha 3$ (S242 with E337

and K319 with E332). **c**, Analysis of the S1–RNB interface. S1 is part of a large hydrophobic interface with RNB ($\alpha 11$, $\alpha 13$ and $\alpha 17$, $\beta 18$ and $\beta 19$). The S1 domain also forms an interface with the RNB through interactions with U8 (also see Fig. 3d). **d**, CSD1 is further stabilized through an RNA-mediated interaction with S1. The backbone phosphate of U4 bridges K78 of CSD1 and R792 of S1.



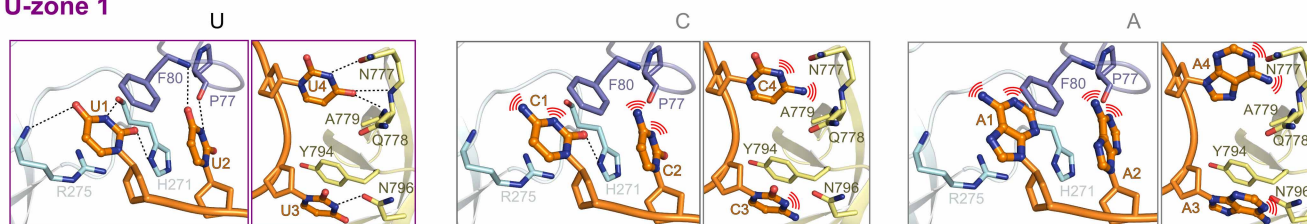
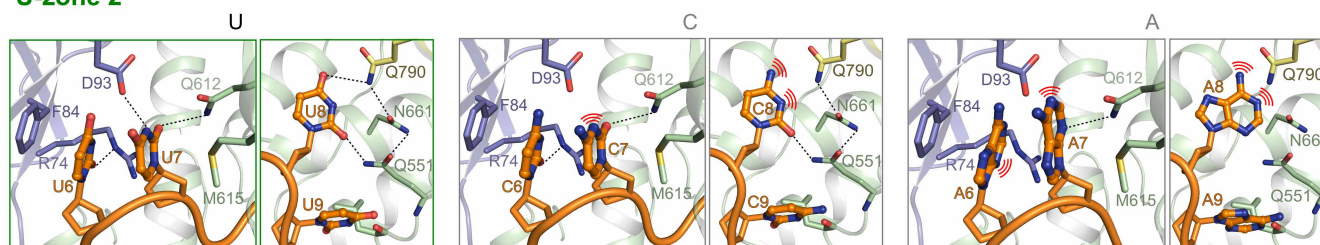
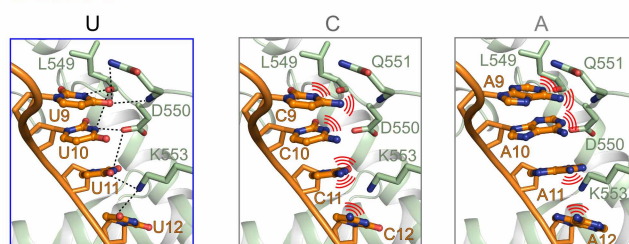
Extended Data Figure 5 | OligoU length preference of Dis3l2. **a**, Specific activity (nt min⁻¹ molecule⁻¹) of Dis3l2 with increasing length oligoU (U₉, U₁₁, U₁₃ and U₁₅). The calculation of specific activity was conducted as in Fig. 1a. Mean ± s.d. ($n = 3$) are shown. **b**, The initial rate plots for each substrate used to calculate specific activity in panel **a**. The amount of substrate degraded (fmol) was plotted against time and the initial rate (slope) was determined with linear regression using GraphPad Prism 6. Mean ± s.d.

($n = 3$) are shown. **c**, Equilibrium dissociation constants (K_d) for increasing oligoU length. The $K_d \pm$ s.d. determined from three independent replicates is shown. **d**, The equilibrium dissociation constant (K_d) for each substrate in panel **c** was determined by plotting the fraction RNA bound versus the concentration of free Dis3l2 and fit by nonlinear regression with GraphPad Prism 6. The mean of three independently measured replicates with error bars representing ± s.d. are shown with the K_d .



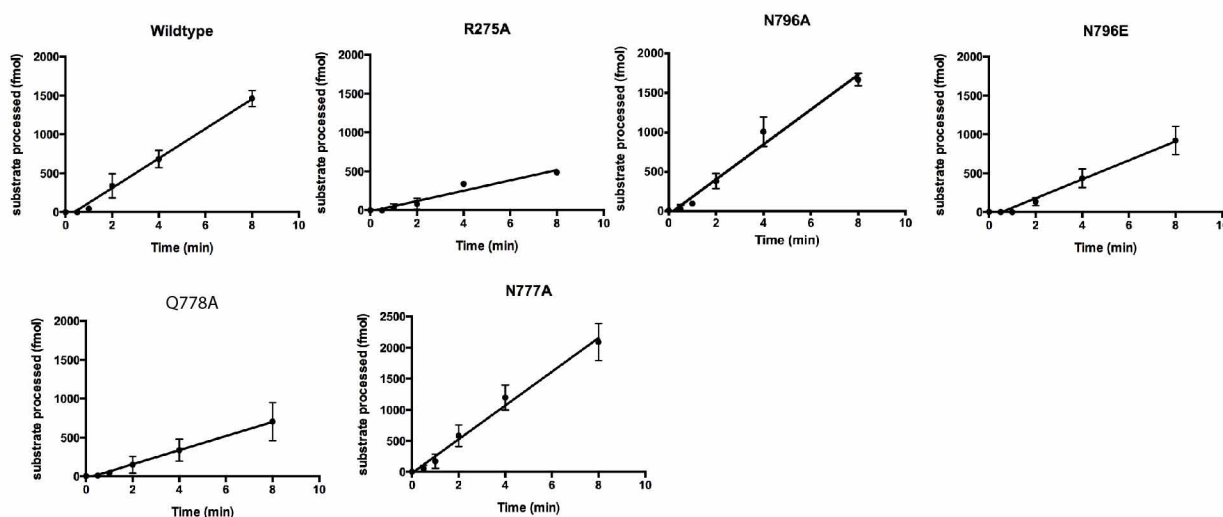
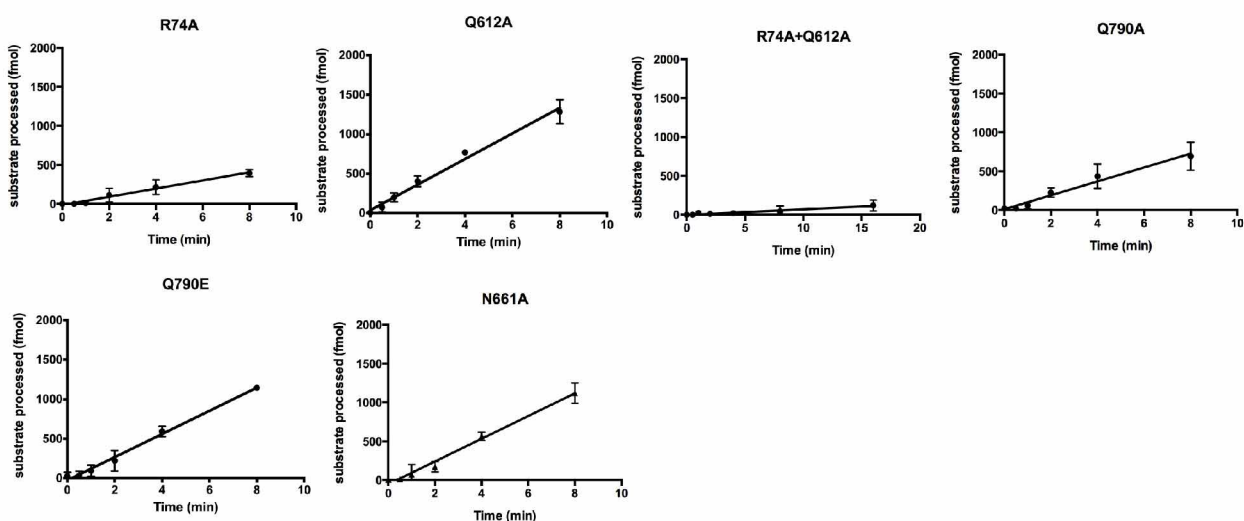
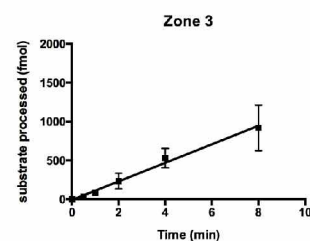
Extended Data Figure 6 | Dis3l2 has a conserved active site and an RNA path that resembles RNase II. **a**, The active site of the Dis3l2-U₁₄ complex. Conserved active site residues are shown as green sticks and U₁₃ and U₁₄ are shown as orange sticks. A single Mg²⁺ ion (purple sphere) is modelled in the active site. As proposed for RNase II, Dis3l2 may utilize a two Mg²⁺ ion mechanism during catalysis. **b**, Superposition of Dis3l2 and yeast Rrp44 (PDB 2VNU) active sites. Dis3l2 side chains and the Mg²⁺ ion are coloured the same as in panel **a**, and yeast Rrp44 residues are shown as grey sticks. **c**, Dis3l2-U₁₄ complex in an identical layout as Fig. 2a. A wide-open funnel created by the CSD lobe and S1 allows RNA to access the top of the RNB. **d**, The structure

of *E. coli* RNase II²² (PDB 2IX1). The path of RNA in RNase II more closely resembles that in Dis3l2, compared to yeast Rrp44, although narrow along its length, underscoring its ability to accommodate only single-stranded RNA substrates. **e**, Perpendicular side and top views of the electrostatic surface potential of Dis3l2 (contoured at $\pm 5 \text{ kT } e^{-1}$; white is neutral, blue is positive and red is negative). A much wider positively charged funnel on the top of the RNB supports the ability of Dis3l2 to degrade structured RNA substrates. **f**, Electrostatic surface potential of the single-strand-specific RNase II, in the same configuration as panel **c**. A narrow RNA binding channel can only accommodate single-stranded RNA.

a U-zone 1**b U-zone 2****c U-zone 3**

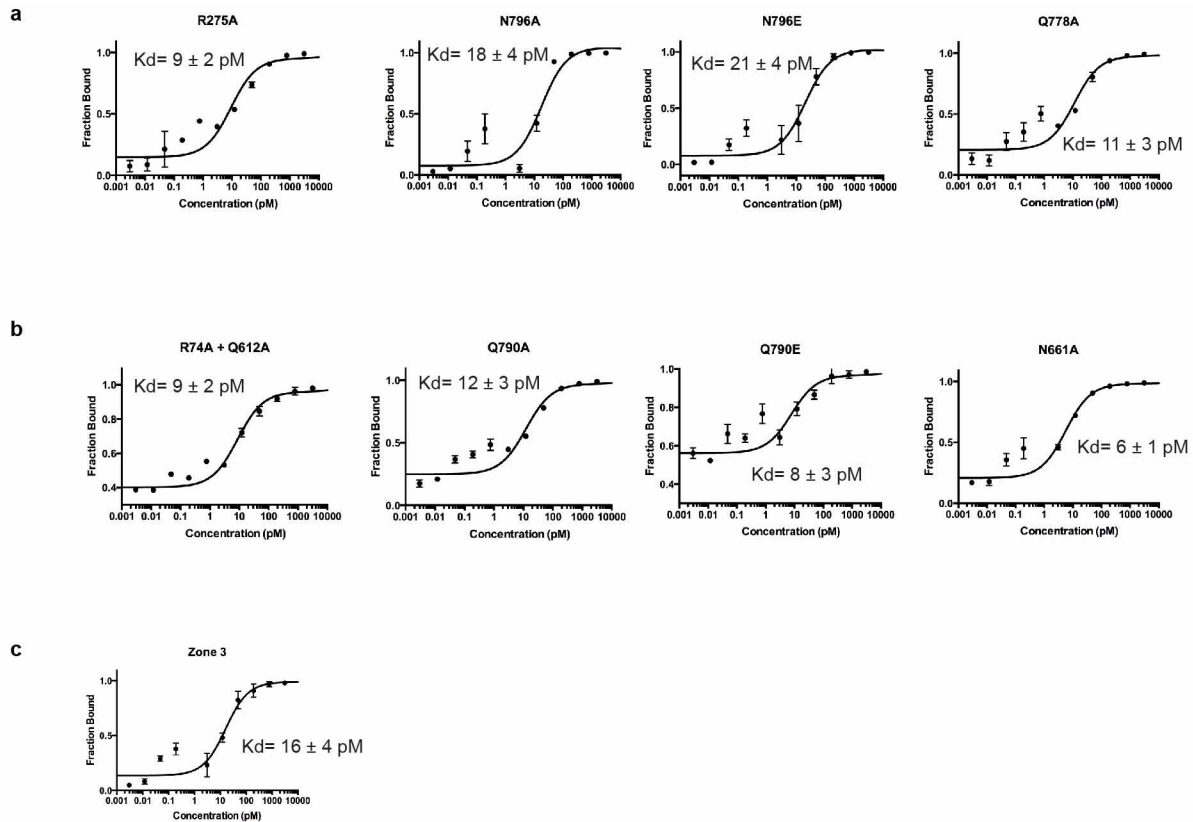
Extended Data Figure 7 | Dis3l2 discriminates U over A and C. **a**, Dis3l2 selects oligoU-tailed substrates by way of an extensive U-specific interaction network along most of its binding path. In U-zone 1, Dis3l2 makes both main-chain and side-chain-mediated hydrogen bonds with uracil. The hydrogen bond network with U₁₄ is disrupted for modelled A₁₄ and C₁₄ RNAs.

Dis3l2 residues are shown as sticks colour-coded by domain as in Fig. 1c. RNA is shown as orange sticks. Hydrogen bond pairs are shown as dashed lines and disrupted hydrogen bonds are shown as curved red lines. **b**, U-zone 2, with the same layout as panel **a**. **c**, U-zone 3.

a**b****c****Extended Data Figure 8 | Initial rate plots of Dis3l2 U-zone mutants.**

a, Initial rate plots for U-zone 1 used to calculate specific activity as shown in Fig. 3f. Assays were conducted under the same conditions as described for U₁₅ in Extended Data Fig. 1, where reactions contained 0.5 nM enzyme and

100 nM 5' radio-labelled U₁₅. Mean \pm s.d. ($n = 3$) are shown. **b**, Initial rate plots for U-zone 2, conducted as described for panel **a**. **c**, Initial rate plots for U-zone 3, conducted as described for panel **a**.



Extended Data Figure 9 | Binding affinity of Dis3l2 U-zone mutants **a**, Slot blot filter-binding assay was used to measure the binding affinity of selected Dis3l2 mutants with U₁₅. The equilibrium dissociation constant (K_d) for U-zone 1 mutants as shown in Fig. 3g was determined by plotting the fraction RNA bound versus the concentration of free protein and fit by nonlinear

regression with GraphPad Prism 6. The mean of three independently measured replicates with error bars representing \pm s.d. are shown with the K_d . **b**, The equilibrium dissociation constant (K_d) for U-zone 2 mutants as shown in Fig. 3g, and measured as described for panel **a**. **c**, The K_d for U-zone 3 as shown in Fig. 3g, and measured as described for panel **a**.

Extended Data Table 1 | Data collection and refinement statistics

Data collection	
Space group	$P2_1$
Cell dimensions	
a, b, c (Å)	63.9 96.1 157.3
α, β, γ (°)	90 98.8 90
Wavelength (Å)	1.1
Resolution (Å)	63.18 - 2.95 (3.06 - 2.95)
R_{merge} (%)	0.128 (0.636)
$I/\sigma(I)$	8.4 (2.1)
Completeness (%)	99.8 (99.9)
Redundancy	3.5 (3.5)
Refinement	
Resolution (Å)	63.2 - 2.95 (3.02 - 2.95)
No. reflections	39755 (3968)
$R_{\text{work}} / R_{\text{free}}$ (%)	0.202/0.251
No. atoms (non-hydrogen)	22942
Protein	22092
RNA	842
Mg^{2+}	2
Water	6
$\langle B\text{-factors} \rangle$ (Å ²)	
Protein	78.3
RNA	101.9
Mg^{2+}	46.6
Water	34.2
r.m.s. deviations	
Bond lengths (Å)	0.012
Bond angles (°)	0.95
Ramachandran	
Favored (%)	96.8
Allowed (%)	3.13
Outliers (%)	0.07

Values in parentheses are for the highest resolution shell.

Required enhancer–matrin-3 network interactions for a homeodomain transcription program

Dorota Skowronska-Krawczyk¹, Qi Ma¹, Michal Schwartz^{1,2}, Kathleen Scully¹, Wenbo Li¹, Zhijie Liu¹, Havilah Taylor¹, Jessica Tollkuhn¹, Kenneth A. Ohgi¹, Dimple Notani¹, Yoshinori Kohwi³, Terumi Kohwi-Shigematsu³ & Michael G. Rosenfeld¹

Homeodomain proteins, described 30 years ago^{1,2}, exert essential roles in development as regulators of target gene expression^{3,4}; however, the molecular mechanisms underlying transcriptional activity of homeodomain factors remain poorly understood. Here investigation of a developmentally required POU-homeodomain transcription factor, Pit1 (also known as Pou1f1), has revealed that, unexpectedly, binding of Pit1-occupied enhancers⁵ to a nuclear matrin-3-rich network/architecture^{6,7} is a key event in effective activation of the Pit1-regulated enhancer/coding gene transcriptional program. Pit1 association with Satb1 (ref. 8) and β -catenin is required for this tethering event. A naturally occurring, dominant negative, point mutation in human PIT1(R271W), causing combined pituitary hormone deficiency⁹, results in loss of Pit1 association with β -catenin and Satb1 and therefore the matrin-3-rich network, blocking Pit1-dependent enhancer/coding target gene activation. This defective activation can be rescued by artificial tethering of the mutant R271W Pit1 protein to the matrin-3 network, bypassing the pre-requisite association with β -catenin and Satb1 otherwise required. The matrin-3 network-tethered R271W Pit1 mutant, but not the untethered protein, restores Pit1-dependent activation of the enhancers and recruitment of co-activators, exemplified by p300, causing both enhancer RNA transcription and target gene activation. These studies have thus revealed an unanticipated homeodomain factor/ β -catenin/Satb1-dependent localization of target gene regulatory enhancer regions to a subnuclear architectural structure that serves as an underlying mechanism by which an enhancer-bound homeodomain factor effectively activates developmental gene transcriptional programs.

During pituitary development, the POU-homeodomain transcription factor, Pit1, is necessary for differentiation of thyrotrope, lactotrope and somatotrope cell types in both mice and humans^{10,11}. To further understand the molecular basis for Pit1-mediated gene activation, we mapped the genomic localization of Pit1 by chromatin immunoprecipitation followed by sequencing (ChIP-seq) using antibody specific against Pit1 (Extended Data Fig. 1a, b) in a growth hormone (GH)-expressing rat pituitary cell line (GC). Of 14,466 Pit1 binding sites identified, >80% overlapped with H3K4me2 histone marks but not with transcription start sites, indicative of enhancer elements⁵ (Fig. 1a, b, Supplementary Table 1).

To identify Pit1 interacting factors, we immunoprecipitated haemagglutinin (HA)-tagged Pit1 from 293T cells, and using mass spectrometry, identified matrin-3 and heterogeneous nuclear ribonucleoprotein U (HNRNPU) as two highly abundant proteins interacting with Pit1 (Fig. 1c, Supplementary Tables 2, 3). The interactions were confirmed by co-immunoprecipitation/western blot experiments in 293T cells (Extended Data Fig. 1c) and with endogenous proteins in GC cells (Fig. 1d, Extended Data Fig. 1d). Both matrin-3 and HNRNPU have been previously found in a salt extraction-resistant nuclear fraction, as part of an internal fibrogranular network^{6,12,13} which we term the nuclear matrin-3-rich network (Extended Data Fig. 1e). These proteins bind DNA at sites known as matrix/scaffold attachment regions¹³.

ChIP-seq using a specific anti-matrin-3 antibody revealed that >50% of matrin-3 binding sites co-localized with H3K4me2 peaks (Extended Data Fig. 1f), showing significant enrichment (~ 16 -fold, P value < 10^{-100}) over the $\sim 3\%$ predicted random co-localization with H3K4me2 peaks, covering $\sim 5\%$ of the rat genome. Of these matrin-3/H3K4me2

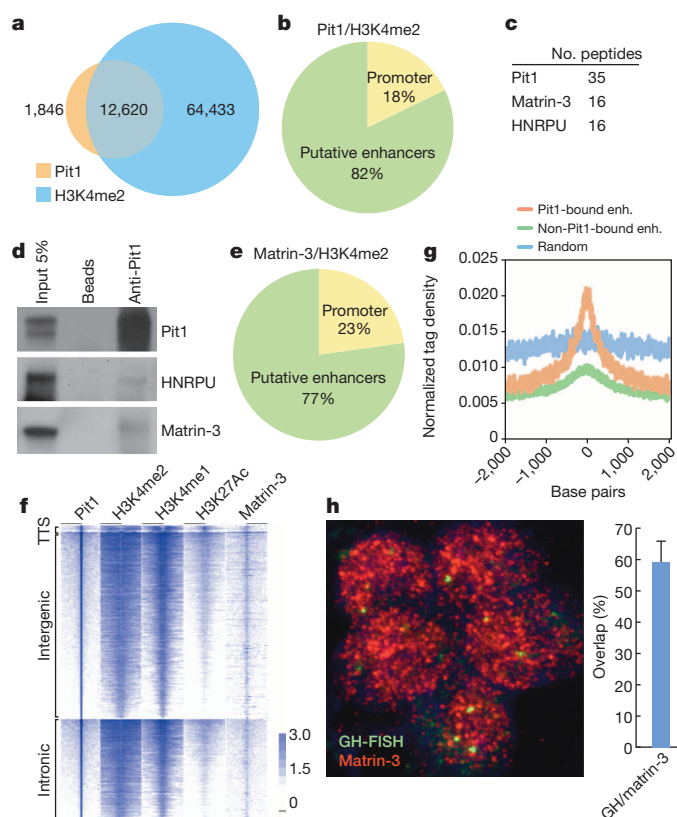


Figure 1 | Co-localization of Pit1 and matrin-3 on enhancers. **a**, ChIP-seq analysis of Pit1 and H3K4me2 genome-wide co-localization in GC rat pituitary somatotrope cell line. **b**, Most DNA sites co-bound by Pit1 and H3K4me2 lie outside gene proximal promoters. **c**, HA-tagged Pit1 immunoprecipitated from 293T cells. Co-purifying factors identified with mass spectrometry. **d**, Co-immunoprecipitation of Pit1 followed by western blot analysis confirmed Pit1: endogenous matrin-3 and HNRNPU interactions in GC cells. **e**, Most matrin-3/H3K4me2 sites in GC cells lie outside gene proximal promoters. **f**, Heat map of ChIP-seq data on non-promoter genome-wide association of Pit1, H3K4me2, H3K4me1, H3K27Ac, matrin-3 in GC cells centred on Pit1 sites and categorized as transcription termination sites (TTS), intergenic, intronic sites. **g**, Meta-analysis of matrin-3 ChIP-seq data. **h**, Example of immuno-FISH experiments showing matrin-3 spots (red) colocalized with GH locus spots (green) in GH-expressing GC cells. Chart represents count of percentage of signals exhibiting co-localization, $n \geq 200$, \pm s.d.

¹Howard Hughes Medical Institute, Department of Medicine, School of Medicine, University of California, San Diego, La Jolla, California 92093, USA. ²The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan, 5290002, Israel. ³Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.

sites, ~80% were elements distal to transcription start sites (Fig. 1e). This finding provides an initial demonstration of matrin-3 association with DNA regulatory sequences in the genome. We noticed that Pit1-associated H3K4me2, H3K4me1 and H3K27Ac-marked enhancers, co-localized with matrin-3 (Fig. 1f, g, Extended Data Fig. 1g), and confirmed matrin-3 antibody specificity by knockdown of endogenous matrin-3 protein (Extended Data Fig. 1h, i). To investigate whether regulatory elements bound by Pit1 occupy the same nuclear compartment as matrin-3, we performed immuno-fluorescence *in situ* (FISH) experiments with DNA-FISH probes specific to the growth hormone (*GH*) locus and anti-matrin-3 antibodies, finding that >65% of the FISH signals co-localized with matrin-3 (Fig. 1h, Extended Data Fig. 1j).

Because the pituitary specific homeodomain protein, Prop1, interacts with β -catenin to activate target gene expression¹⁴, we investigated β -catenin interacting partners in differentiated pituitary cells by expressing β -catenin fused to the biotin ligase recognition peptide (BLRP), along with biotinylating enzyme BirA, in GC cells. Biotinylated β -catenin was present in both the cytoplasmic and the nuclear compartments (Fig. 2a, Extended Data Fig. 2a). Streptavidin pull-down, followed by mass spectrometry, revealed that whereas cytoplasmic β -catenin-interacting factors included α -catenin, as previously described¹⁵, Pit1 peptides predominated in the nuclear β -catenin fraction (Fig. 2a, Supplementary Tables 4–7). This putative Pit1– β -catenin interaction was confirmed between endogenous proteins by co-immunoprecipitation/western blot experiments using nuclear extracts from GC pituitary cells (Fig. 2b). Using glutathione-S-transferase (GST) pull-down assays,

armadillo repeats 8–12 of β -catenin proved sufficient to mediate this interaction, exactly within a region that has been previously described to interact with Prop1 (ref. 14) as well as Lef/Tcf¹⁶ (Fig. 2c, Extended Data Fig. 2b).

We then examined whether β -catenin plays a role in regulating expression of Pit1 target genes, such as *GH*. Effective knockdown of β -catenin in GC cells (Extended Data Fig. 2c) markedly decreased the level of nascent *GH* transcripts (Fig. 2d), analogous to what was observed when Pit1 was knocked down (Fig. 2d, Extended Data Fig. 2c). β -catenin also forms a complex with Satb1 (ref. 17), a well-established genome organizer essential for multiple biological processes, including T-cell activation and cancer progression^{8,18,19}. Reverse transcription coupled with quantitative PCR (RT-qPCR) analysis revealed that the upregulation of Satb1 expression in the mouse pituitary gland coincides with differentiation of Pit1 lineages during pituitary development (Extended Data Fig. 2d), and immunohistochemical analysis confirmed that SATB1 protein is expressed in somatotropes and other cells of the mouse pituitary gland (Extended Data Fig. 2e). Co-immunoprecipitation of Pit1 from GC cells revealed an interaction with Satb1 (Fig. 2b) which was reproduced by co-immunoprecipitation of HA-tagged Pit1 in 293T cells (Extended Data Fig. 2f). Indeed, conditional knockout of *Satb1* in the mouse anterior pituitary significantly decreased expression of *GH* (Extended Data Fig. 2g). We therefore investigated whether interaction between Pit1 and Satb1 was dependent on β -catenin, and vice-versa. Exogenously expressed, HA-tagged Pit1 protein was pulled down after knocking down either β -catenin or Satb1 in 293T cells. In the absence of β -catenin, the

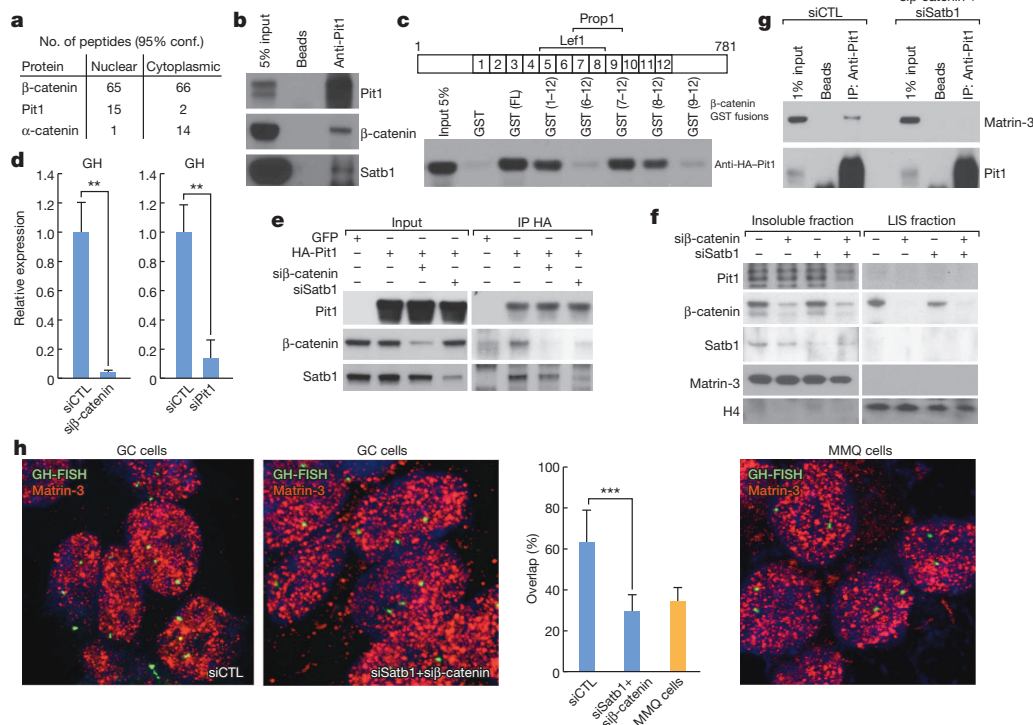


Figure 2 | Pit1 association with LIS resistant nuclear component is β -catenin- and SATB1-dependent. **a**, BLRP- β -catenin immunoprecipitated from cytoplasmic or nuclear fractions of GC cells. Co-purifying factors identified by mass spectrometry. **b**, Co-immunoprecipitation-Western analysis confirmed Pit1: β -catenin interaction and revealed interaction with SATB1. **c**, GST-pull-down showing β -catenin armadillo repeat 8 (within region previously shown to interact with Prop1 and Lef1) is required for interaction with Pit1. **d**, Nascent GH transcripts levels after siRNA knockdown of β -catenin and Pit1 analysed by RT-qPCR. Experiments were repeated 2 times, and *P* values calculated using Student's two tailed *t*-test (\pm s.d.; $**P < 0.01$). **e**, Co-immunoprecipitation of HA-tagged Pit1 protein in 293T cells before and after β -catenin and Satb1 knockdown showing Pit1 interacts simultaneously

with both proteins. **f**, LIS nuclear extraction before and after β -catenin or/and Satb1 knockdown shows that both proteins are needed for Pit1 retention in LIS resistant fraction. **g**, Co-immunoprecipitation of Pit1 protein in GC cells before and after simultaneous knockdown of β -catenin and Satb1 shows interaction of Pit1 with matrin-3 is dependent on the presence of both proteins. **h**, Examples of immuno-FISH experiments showing GH locus (green) colocalizing with matrin-3 (red) with higher frequency in control culture conditions than after siRNA β -catenin and siRNA Satb1 treatment in GC cells. In Pit1 positive, non-GH-expressing MMQ cells, significantly fewer GH loci are associated with matrin-3. The chart represents the count of percentage of signals exhibiting co-localization in control vs. si β -catenin and siSATB1 conditions compared to MMQ cells; $n \geq 200$, \pm s.d., $***P < 0.001$.

interaction between Pit1 with Satb1 was diminished, indicating a role for β -catenin in mediating Pit1–Satb1 interaction. Reciprocally, the interaction of Pit1 with β -catenin was partially abolished by Satb1 knock-down (Fig. 2e).

Pit1 has been shown previously to co-fractionate with the insoluble, matrin-3 rich, nuclear fraction in *in-vitro* biochemical studies²⁰. We therefore used lithium 3,5-diiodosalicylate (LIS), a chaotropic reagent that extracts a majority of nuclear proteins but preserves the nuclear matrin-3-rich network²¹, to investigate Pit1 subnuclear locations (Extended Data Fig. 2h). We observed that, in somatotropes, Pit1 indeed co-fractionated with matrin-3 and Satb1 in the LIS resistant insoluble fraction, while β -catenin was partitioned between the insoluble and soluble fractions (Extended Data Fig. 2i). When both β -catenin and Satb1 proteins were simultaneously depleted (Extended Data Fig. 2j), the presence of Pit1 in the LIS resistant insoluble fraction was significantly reduced (Fig. 2f). In the double knockdown, total Pit1 protein levels were unchanged and Pit1 protein remained localized to the nucleus (Extended Data Fig. 2k); however, it was now detectable in the looped-out DNA fraction obtained

after DNAse I digestion of nuclear halos (Extended Data Fig. 2l). Together, these data indicate that localization of Pit1 protein to the insoluble fraction is dependent on interaction with β -catenin and Satb1. Additionally, when endogenous Pit1 protein was pulled down before and after knocking down both β -catenin and Satb1 in GC cells, the absence of these proteins highly affected interaction between Pit1 and matrin-3 (Fig. 2g).

To further investigate whether association of Pit1-bound enhancers with the matrin-3 fraction is indeed dependent on β -catenin and SATB1, we performed immuno-FISH experiments in GC cells with anti-matrin-3 antibodies and DNA-FISH probes specific to the growth hormone (GH) locus. In GH-expressing GC cells, ~65% of the FISH signals co-localized with matrin-3. When both β -catenin and Satb1 were depleted, only ~35% of FISH signals loci co-localized with matrin-3, similar to the overlap of a GH locus in a Pit-1 positive, GH-non-expressing cell line (MMQ) (Fig. 2h, Extended Data Fig. 2m). We thus conclude that association of Pit1 enhancers with the nuclear matrin-3-rich network/structure observed by immuno-FISH (Fig. 1h, Fig. 2h) and biochemical extractions (Fig. 2f, Extended Data Fig. 2h) is dependent on intact interactions of Pit1 with both β -catenin and Satb1.

Global run-on (GRO-seq) analysis using GC cells before and after simultaneous knockdown of both β -catenin and Satb1 revealed that in total, 1,350 coding gene nascent transcripts were downregulated and 916 genes were upregulated under these double-knockdown conditions (Extended Data Fig. 3a). Out of the 1,350 genes positively regulated by both β -catenin and SATB1 (Supplementary Table 8), 991 had Pit1-bound enhancer elements within 200 kbp of their start sites (Extended Data Fig. 3a), and all 991 were significantly downregulated (Fig. 3a). Consistent with their functional importance²² the expression of the eRNAs associated with these 991 Pit1-bound enhancers was significantly affected in the simultaneous knockdown of β -catenin and Satb1 (Fig. 3b). ChIP-seq analysis revealed that Satb1 and β -catenin proteins were located at the centre of Pit1-positive enhancers (Extended Data Fig. 3b–d), strongly supporting the co-binding model of these three proteins. Based on this analysis, we selected several highly downregulated genes that contain Pit1 enhancers (Extended Data Fig. 3c) and observed that knockdown of either of the two proteins significantly diminished target gene expression (Extended Data Fig. 3g). Additionally, knockdown of either Pit1 or β -catenin affected Satb1 association with enhancer elements, as assessed by ChIP-qPCR (Fig. 3c, Extended Data Fig. 3e). Association of matrin-3 with Pit1-dependent enhancers was significantly reduced when either β -catenin or SATB1 were downregulated (Extended Data Fig. 3h). Finally, we confirmed that lack of Pit1 protein inhibited association of selected Pit1-dependent enhancers with matrin-3 (Fig. 3d, Extended Data Fig. 3f). Together, these data suggest that functional interaction between Pit1-dependent regulatory elements and matrin-3 requires association of the DNA-bound Pit1 with β -catenin and SATB1. Consistent with this suggestion, lack of matrin-3 had a negative effect on the expression of target genes (Extended Data Fig. 3i).

Naturally occurring mutations in Pit1 underlie number of combined pituitary hormone deficiency (CPHD) cases in humans, with the R271 residue being the most frequently mutated⁹. Amino acid 271 maps to the last turn of helix 3 in the homeodomain and is located outside the DNA-binding surface. Although the R271W mutation in Pit1 confers dominant-negative properties (Extended Data Fig. 4a)⁹, it does not affect protein stability or binding to its cognate DNA sites (Extended Data Fig. 4b–d)⁹. We therefore investigated whether the R271W mutation affects the composition of the Pit1 complex. Co-immunoprecipitation of recombinant HA-tagged Pit1 proteins followed by western blot analysis revealed that the R271W mutation disrupted the ability of Pit1 to interact with matrin-3 and HNRNPU, apparently owing to its inability to bind β -catenin and Satb1 (Fig. 4a). Additionally, biochemical fractionation indicated that the R271W mutant is largely lost from the insoluble fraction (Fig. 4b), although both wild-type and R271W proteins were expressed at similar levels and found entirely localized to the nucleus (Extended Data Fig. 4e). Instead, R271W was detected in the

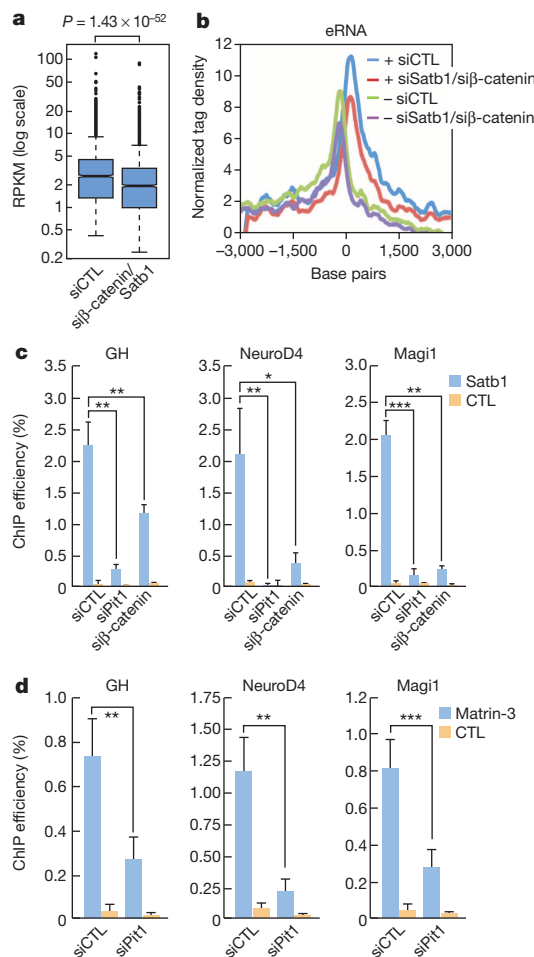


Figure 3 | β -catenin and SATB1 influence transcription of genes regulated by Pit1 enhancers. **a**, Genes containing Pit1 enhancers (991) are significantly downregulated upon knockdown of β -catenin and Satb1. RPKM – reads per kilobase per million. **b**, GRO-seq tag distribution on both strands of Pit1 bound enhancers before and after treatment of GC cells with β -catenin and Satb1 siRNAs. **c**, ChIP-qPCR analysis of Satb1 association with Pit1 enhancers upon either Pit1 or β -catenin knockdown. Experiments were repeated 3 times, and *P* values calculated using Student's two tailed *t*-test (\pm s.d.; **P* < 0.05, ***P* < 0.01, ****P* < 0.001). **d**, ChIP-qPCR analysis of matrin-3 association with Pit1 enhancers upon Pit1 knockdown. Experiments repeated 2–4 times, *P* values calculated using Student's two tailed *t*-test (\pm s.d.; **P* < 0.05, ***P* < 0.01, ****P* < 0.001).

'looped-out DNA' fraction (Fig. 4b), suggesting that it had lost its ability to interact with the LIS-resistant fraction.

Together, these data indicated that a key function of homeodomain transcription factors such as Pit1 might be to join bound regulatory regions to the matrin-3-rich network, an association required to activate target gene transcription. Because the transcriptionally inactive R271W mutation interferes with this association, we experimentally tested this hypothesis by artificially tethering the R271W mutant to the matrin-3 network to try to restore target gene transcription. Several protein domains that are responsible for association with the matrin-3-rich 'matrix attachment regions' have been previously described and include the SAF/SAP domain (for example, in HNRNP23, SAF-B, PARP proteins²⁴); AT-hooks (HMG proteins²⁵); and the atypical homeodomain, SATB1¹⁸. We therefore grafted the SAF/SAP domain from rat HNRNP23 (amino acids 4–45), onto HA-tagged wild-type and R271W Pit1 (Fig. 4c). As predicted, R271W-SAF hybrid protein localized efficiently in the insoluble nuclear fraction (Extended Data Fig. 4f). Hybrid wild-type (WT)-SAF and R271W-SAF proteins were expressed in GC cells treated with short interfering RNAs to deplete endogenous Pit1 (Extended Data Fig. 4g). As expected, matrin-3 association with Pit1-dependent enhancers and expression of Pit1-dependent coding genes and eRNAs were lost upon

Pit1 knockdown, but were restored in the presence of either wild-type or WT-SAF Pit1 (Fig. 4d, Extended Data Fig. 4i). In contrast, neither the association of Pit1 enhancers with matrin-3 nor target gene and eRNA expression were rescued in R271W-expressing cells. However, the R271W-SAF hybrid rescued matrin-3 interaction and Pit1 transcriptional activation to levels close to those achieved by wild-type Pit1. These experiments indicate that specific subnuclear localization of Pit1 is a bona fide requirement for its full activity, and that this is the only functional defect in the R271W mutant Pit1 protein.

To investigate which step of enhancer activation is affected by R271W, we tested recruitment of p300, functional homologue of CREB binding protein (CBP), a known co-factor of Pit1²⁶, finding that p300 association with tested enhancer elements was significantly reduced (Extended Data Fig. 4h). Overexpression of R271W was not able to rescue this recruitment whereas hybrid R271W-SAF did restore p300 association with Pit1 bound enhancers (Extended Data Fig. 4h). These data suggest that eRNA transcription and binding of critical co-activators, such as p300, require recruitment of Pit1 to the matrin-3-rich network.

To further substantiate these observations in intact pituitary, *ex vivo* rescue approach took advantage of the "Snell" mouse model that harbours a mutation (W261C) in Pit1 that renders it incapable of binding

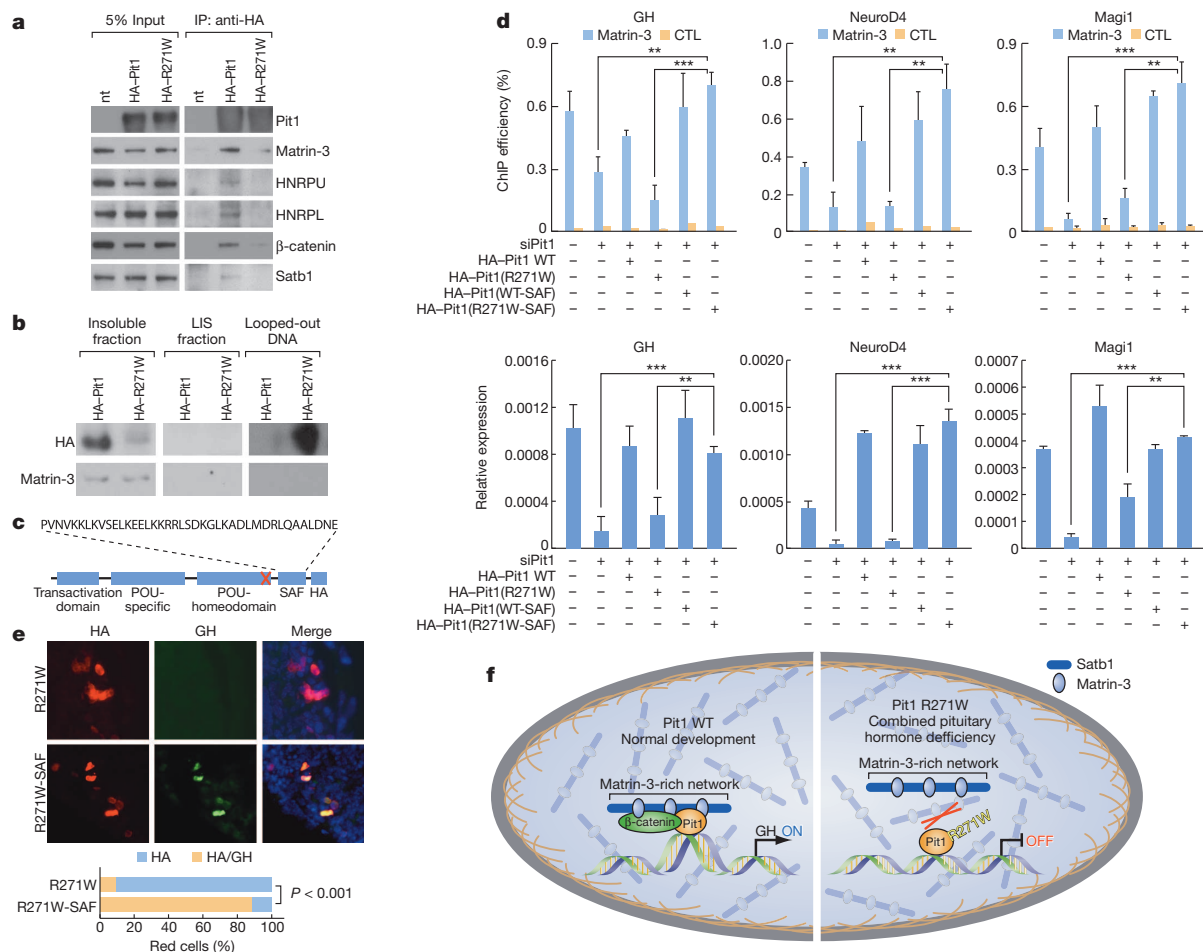


Figure 4 | Naturally occurring R271W mutation in human Pit1 affects ability of Pit1 to interact with matrin-3 rich network. **a**, Western blot analysis showing defective interaction of Pit1R271W mutant with components of nuclear matrin-3 enriched network as well as β -catenin and Satb1. **b**, Biochemical extraction shows that the amount of R271W mutant protein decreased in the LIS-resistant fraction and associates instead with the looped-out DNA fraction. **c**, Schematic representation of Pit1 protein with added SAF domain. Red "x" – location of amino-acid 271. **d**, Analysis of effect of overexpression of different forms of Pit1 protein in absence of endogenous Pit1; top panels: ChIP-qPCR analysis of matrin-3 association; bottom panels: RT-qPCR analysis of transcriptional effect on selected targets. Experiments

were repeated 3 or 4 times, and *P* values calculated using Student's two tailed *t*-test (\pm s.d.; ***P* < 0.01, ****P* < 0.001). **e**, Analysis of effect of *ex vivo* overexpression of R271W mutant with and without SAF domain in P2 *snell* mutant pituitary glands. GH expression is rescued only when the R271W-SAF protein is overexpressed. Bottom: quantification of percentage of HA-positive GH expressing cells. Two pituitaries were used per condition; *P* value calculated using χ^2 test. **f**, Model: association of Pit1-bound regulatory element with matrin-3 rich network and loss of this interaction based on failure of R271W mutant to interact with β -catenin and SATB1 causing loss of gene activation as in CPHD.

to cognate DNA sites. Overexpressing HA-tagged Pit1-R271W-SAF hybrid, but not Pit1-R271W protein, rescued expression of endogenous GH as visualized by immunohistochemistry (Fig. 4e). On the basis of these data, we propose a model in which Pit1 association with the matrin-3-rich network is indispensable for its biological activity, as demonstrated by the lack of transcriptional activity of the naturally occurring mutant R271W Pit1 (Fig. 4f).

Nuclear organization and genome three dimensional structure are postulated to have a prominent role in regulation of gene expression^{28,29}. Here, we have presented evidence that a homeodomain transcription factor regulates gene transcriptional programs through interaction with components of subnuclear structure (for example, matrin-3-rich structure). Other homeodomain proteins (for example, Lhx3, Prop1, Oct1), as well as different classes of transcription factors³⁰, have been shown to associate with the high-salt extraction-resistant fraction, corresponding to the matrin-3-rich network. This suggests a model in which this interaction may be required for transcriptional activation of enhancers regulated by many homeodomain proteins, and perhaps several other types of DNA binding transcription factors.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 2 December 2013; accepted 10 June 2014.

Published online 3 August 2014.

- McGinnis, W., Levine, M. S., Hafen, E., Kuroiwa, A. & Gehring, W. J. A conserved DNA sequence in homoeotic genes of the *Drosophila* Antennapedia and bithorax complexes. *Nature* **308**, 428–433 (1984).
- Scott, M. P. & Weiner, A. J. Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu loci of *Drosophila*. *Proc. Natl Acad. Sci. USA* **81**, 4115–4119 (1984).
- Affolter, M., Schier, A. & Gehring, W. J. Homeodomain proteins and the regulation of gene expression. *Curr. Opin. Cell Biol.* **2**, 485–495 (1990).
- Prince, K. L., Walvoord, E. C. & Rhodes, S. J. The role of homeodomain transcription factors in heritable pituitary disease. *Nature Rev. Endocrinol.* **7**, 727–737 (2011).
- Heintzman, N. D. & Ren, B. Finding distal regulatory elements in the human genome. *Curr. Opin. Genet. Dev.* **19**, 541–549 (2009).
- Zeit, M. J., Malyavantham, K. S., Seifert, B. & Berezney, R. Matrin 3: chromosomal distribution and protein interactions. *J. Cell. Biochem.* **108**, 125–133 (2009).
- Nickerson, J. Experimental observations of a nuclear matrix. *J. Cell Sci.* **114**, 463–474 (2001).
- Kohwi-Shigematsu, T. et al. Genome organizing function of SATB1 in tumor progression. *Semin. Cancer Biol.* **23**, 72–79 (2013).
- Radovick, S. et al. A mutation in the POU-homeodomain of Pit-1 responsible for combined pituitary hormone deficiency. *Science* **257**, 1115–1118 (1992).
- Zhu, X., Wang, J., Ju, B. G. & Rosenfeld, M. G. Signaling and epigenetic regulation of pituitary development. *Curr. Opin. Cell Biol.* **19**, 605–611 (2007).
- Ingraham, H. A. et al. A tissue-specific transcription factor containing a homeodomain specifies a pituitary phenotype. *Cell* **55**, 519–529 (1988).
- Nakayasu, H. & Berezney, R. Nuclear matrices: identification of the major nuclear matrix proteins. *Proc. Natl Acad. Sci. USA* **88**, 10312–10316 (1991).
- Romig, H., Fackelmayer, F. O., Renz, A., Ramsperger, U. & Richter, A. Characterization of SAF-A, a novel nuclear DNA binding protein from HeLa cells with high affinity for nuclear matrix/scaffold attachment DNA elements. *EMBO J.* **11**, 3431–3440 (1992).
- Olson, L. E. et al. Homeodomain-mediated beta-catenin-dependent switching events dictate cell-lineage determination. *Cell* **125**, 593–605 (2006).
- Su, L. K., Vogelstein, B. & Kinzler, K. W. Association of the APC tumor suppressor protein with catenins. *Science* **262**, 1734–1737 (1993).
- Graham, T. A., Weaver, C., Mao, F., Kimelman, D. & Xu, W. Crystal structure of a beta-catenin/Tcf complex. *Cell* **103**, 885–896 (2000).
- Notani, D. et al. Global regulator SATB1 recruits beta-catenin and regulates T(H)2 differentiation in Wnt-dependent manner. *PLoS Biol.* **8**, e1000296 (2010).
- Dickinson, L. A., Dickinson, C. D. & Kohwi-Shigematsu, T. An atypical homeodomain in SATB1 promotes specific recognition of the key structural element in a matrix attachment region. *J. Biol. Chem.* **272**, 11463–11470 (1997).
- Cai, S., Han, H. J. & Kohwi-Shigematsu, T. Tissue-specific nuclear architecture and gene expression regulated by SATB1. *Nature Genet.* **34**, 42–51 (2003).
- Mancini, M. G., Liu, B., Sharp, Z. D. & Mancini, M. A. Subnuclear partitioning and functional regulation of the Pit-1 transcription factor. *J. Cell. Biochem.* **72**, 322–338 (1999).
- Mirkovitch, J., Mirault, M. E. & Laemmli, U. K. Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold. *Cell* **39**, 223–232 (1984).
- Li, W. et al. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* **498**, 516–520 (2013).
- Kipp, M. et al. SAF-Box, a conserved protein domain that specifically recognizes scaffold attachment region DNA. *Mol. Cell. Biol.* **20**, 7480–7489 (2000).
- Aravind, L. & Koonin, E. V. SAP – a putative DNA-binding motif involved in chromosomal organization. *Trends Biochem. Sci.* **25**, 112–114 (2000).
- Reeves, R. Molecular biology of HMGA proteins: hubs of nuclear function. *Gene* **277**, 63–81 (2001).
- Xu, L. et al. Signal-specific co-activator domain requirements for Pit-1 activation. *Nature* **395**, 301–306 (1998).
- Li, S. et al. Dwarf locus mutants lacking three pituitary cell types result from mutations in the POU-domain gene *pit-1*. *Nature* **347**, 528–533 (1990).
- Gibcus, J. H. & Dekker, J. The hierarchy of the 3D genome. *Mol. Cell* **49**, 773–782 (2013).
- Bickmore, W. A. & van Steensel, B. Genome architecture: domain organization of interphase chromosomes. *Cell* **152**, 1270–1284 (2013).
- S'iakste, N. I. & S'iakste, T. G. Transcription factors and the nuclear matrix [in Russian]. *Mol. Biol. (Mosk.)* **35**, 739–749 (2001).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank M. Ghassemlian for assistance with mass spectrometry; C. Nelson for cell culture assistance; J. Hightower for assistance with figures and manuscript preparation; T. Suter for help with images analysis. We acknowledge J. Santini and the UCSD Neuroscience Microscopy Shared Facility (Grant P30 NS047101) for imaging. These studies were supported by grants NS034934, DK039949, DK018477, HL065445, CA173903 to M.G.R. from NIH. D.S.-K. was supported by EMBO Long Term Fellowship, The Swiss National Science Foundation and The San Diego Foundation. M.G.R. is an Investigator with HHMI.

Author Contributions D.S.-K. and M.G.R. conceived the project. D.S.-K. performed the majority of the experiments; Q.M. performed the bioinformatic analyses; M.S. contributed three dimensional immuno-FISH. W.L. contributed GRO-Seqs; K.A.O. assisted in deep-sequencing library preparations and sequencing; additional experiments/methods were contributed by K.S., J.T., Z.L. and D.N.; H.T. and K.S. assisted in animal-based experiments. Y.K. and T.K.-S. contributed critical insights and reagents. D.S.-K., K.S. and M.G.R. wrote the manuscript. All authors discussed the results and commented on or edited the manuscript.

Author Information Microarray data are deposited in GEO under accession number GSE58009. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.G.R. (mrosenfeld@ucsd.edu) or D.S.-K. (dkrawczyk@ucsd.edu).

METHODS

Antibodies. The antibodies used in this study are listed in Supplementary Table 9. **siRNA and expression vector electroporations.** siRNA and overexpression experiments in GC cells and in P2 pituitary were performed using Neon system (Invitrogen). Conditions of 2 pulses of 1,150 V with 30 ms spacing were used. 2.5 μ M siRNA and 30 μ g of DNA were used per 3×10^6 cells. Knockdown was repeated after 48 h to obtain optimal depletion of the protein, verified by western blotting. Overexpression experiments were harvested after 48 h. siRNAs used in the study are listed in Supplementary Table 10. siRNA and overexpression experiments in 293T cells were performed using Lipofectamine 2000 (Invitrogen) using manufacturer's protocol. **RT-qPCR.** RNA was isolated using TRIzol (Invitrogen) and total RNA was reverse transcribed using SuperScript III Reverse Transcriptase (Invitrogen). Quantitative PCRs were performed mostly with SsoAdvanced SYBR Green Supermix (BIO-RAD). For normalization, ΔC_t values were calculated relative to the levels of *GAPDH* transcripts. The experiments were repeated at least three times as a biological replicates, *P* values were obtained using a two-tailed Student's *t*-test; \pm s.d., **P* < 0.05, ***P* < 0.01, ****P* < 0.001; primers are listed in Supplementary Table 10.

ChIP-seq. ChIP was performed as previously described¹. Briefly, for ChIP-seq, approximately 10^7 cells were cross-linked with 1% formaldehyde at room temperature for 10 min and neutralized with 0.125 M glycine. After sonication, 20–30 μ g of soluble chromatin was incubated with 1–5 μ g of antibody at 4 °C overnight. Immunoprecipitated complexes were collected using G-protein Dynabeads (Invitrogen). Subsequently, immuno-complexes were washed, and DNA was extracted and purified by QIAquick Spin columns (Qiagen). For ChIP-seq, the extracted DNA was ligated to specific adaptors followed by deep sequencing with the Illumina's HiSeq 2000 system according to the manufacturer's instructions. Usually, the first 48 bp for each sequence tag returned by the Illumina Pipeline was aligned to the rn4 assembly using BFAST or Bowtie2. Only uniquely mapped tags were selected for further analysis. The data was visualized by preparing custom tracks on the UCSC genome browser using HOMER (<http://homer.salk.edu/homer>). The total number of mappable reads was normalized to 10^7 for each experiment presented in this study.

Identification of ChIP-seq peaks. The ChIP-seq peaks were identified by HOMER. Given different binding patterns of transcription factors and histones, parameters were optimized for the narrow tag distribution characteristic of transcription factors by searching for high read-enrichment regions within a 200-bp sliding window. Regions of maximal density exceeding a given threshold were called as peaks, and adjacent peaks were set to be >500 bp away to avoid redundant detection. The common artefacts from clonal amplification were circumvented by considering only one tag from each unique genomic position. The threshold was set at a false discovery rate (FDR) of 0.001 determined by peak finding using randomized tag positions in a genome with an effective size of 2×10^9 bp. For ChIP-seq of histone marks, seed regions were initially found using a peak size of 500 bp (FDR < 0.001) to identify enriched loci. Enriched regions separated by <1 kb were merged and considered as blocks of variable lengths. All called peaks were then associated with genes by cross-referencing with the RefSeq TSS database. Peaks from individual experiments were considered overlapping if their peak centres were located within 200 bp (for some analyses the distance between them could extend to 1 kb). The peaks within ± 1 kb apart from the RefSeq gene TSS site were considered to be promoter bound.

GRO-seq. GRO-seq experiments were performed as previously reported^{21,32}. Briefly, GC cells were swelled in swelling buffer (10 mM Tris-Cl pH 7.5, 2 mM MgCl₂, 3 mM CaCl₂) for 5 min on ice and then lysed in lysis buffer (swelling buffer with 0.5% IGEPAL and 10% glycerol) before being finally re-suspended in 100 μ l of freezing buffer (50 mM Tris-Cl pH 8.3, 40% glycerol, 5 mM MgCl₂, 0.1 mM EDTA). For the run-on assay, re-suspended nuclei were mixed with an equal volume of reaction buffer (10 mM Tris-Cl pH 8.0, 5 mM MgCl₂, 1 mM dithiothreitol (DTT), 300 mM KCl, 20 units of Superase-In, 1% sarkosyl, 500 μ M ATP, GTP, Br-UTP and 2 μ M CTP) and incubated for 5 min at 30 °C. The nuclear-run-on RNA (NRO-RNA) was then extracted with TRIzol LS reagent (Invitrogen) following the manufacturer's instructions. After base hydrolysis on ice for 40 min and followed by treatment with DNase I and Antarctic phosphatase, the Br-UTP-labelled NRO-RNA was purified by anti-BrdU agarose beads (Santa Cruz Biotech) in binding buffer (0.5 \times SSPE, 1 mM EDTA, 0.05% Tween) for 3 h at 4 °C while rotating. Then T4 PNK (NEB) was used to repair the end of NRO-RNA. Subsequently, complementary DNA synthesis was performed as reported with few modifications. The RNA fragments were subjected to the poly-A-tailing reaction by poly-A polymerase (NEB) for 30 min at 37 °C. Reverse transcription was then performed using superscript III (Invitrogen) with oNT1223 primer (for sequence see Supplementary Table 10). The cDNA products were separated on a 10% polyacrylamide TBE-urea gel with the right product (~100–500 bp) being excised and recovered by gel extraction. After that, the first-strand cDNA was circularized by CircLigase (Epicentre) and re-linearized by Ape1 (NEB). Re-linearized single-strand cDNA was separated by TBE gel and the products of

the desired size were excised (~120–320 bp) for gel extraction. Finally, the cDNA template was amplified by PCR using the Phusion High-Fidelity enzyme (NEB) with primers oNT1200 and oNT1201 for deep sequencing (primers listed in Supplementary Table 10).

GST-affinity purification and protein interaction studies. GST- β -catenin fusion vectors for protein interactions (gift from C. Neuvet) were expressed in *Escherichia coli* and purified from homogenized lysates with glutathione-agarose beads at 25 °C for 1 h. For interaction studies, similar amounts of immobilized GST-fusion proteins were then mixed with equal amounts of 293T cell lysates containing overexpressed, HA-tagged Pit1 protein and washed 4 \times with 1 \times PBS with 10 mM DTT and protease inhibitors. Interacting proteins were eluted with 20 mM L-glutathione pH 8.0 and visualized by western blotting.

Immunoprecipitation. Cells were collected with cold PBS and lysed with Dignam buffer C (20 mM HEPES pH 7.9, 25% glycerol, 420 mM NaCl, 1.5 mM MgCl₂, 0.2 mM EDTA, 0.5 mM DTT, protease inhibitors). The lysate was diluted with dilution buffer (20 mM HEPES pH 7.9, 1.5 mM MgCl₂, 0.2 mM EDTA, 0.5 mM DTT, protease inhibitors) to the final 150 mM NaCl concentration. 2–5 μ g of antibodies were added into the diluted cell lysate and incubated overnight at 4 °C. The next day, the protein complexes were collected by magnetic Dynabeads G for 2 h at 4 °C with rotation. The beads-antibody-protein complexes were then washed four times with wash buffer (PBS, 0.03% Triton X-100) and boiled for western blot analysis. For cytoplasmic/nuclear analysis and immunoprecipitation followed by liquid chromatography and tandem mass spectrometry (LC-MS-MS) nuclei were isolated (using standard Dignam protocol) before lysis.

Sample preparation for LC-MS-MS. Protein samples were diluted in TNE (50 mM Tris pH 8.0, 100 mM NaCl, 1 mM EDTA) buffer. RapiGest SF reagent (Waters Corp.) was added to the mix to a final concentration of 0.1% and samples were boiled for 5 min. TCEP (Tris (2-carboxyethyl) phosphine) was added to 1 mM (final concentration) and the samples were incubated at 37 °C for 30 min. Subsequently, the samples were carboxymethylated with 0.5 mg ml⁻¹ of iodoacetamide for 30 min at 37 °C followed by neutralization with 2 mM TCEP (final concentration). Proteins samples prepared as above were digested with trypsin (trypsin:protein ratio 1:50) overnight at 37 °C. RapiGest was degraded and removed by treating the samples with 250 mM HCl at 37 °C for 1 h followed by centrifugation at 14,000 r.p.m. for 30 min at 4 °C. The soluble fraction was then added to a new tube and the peptides were extracted and desalted using C18 desalting columns (Thermo Scientific).

LC-MS-MS analysis. Trypsin-digested peptides were analysed by ultrahigh pressure liquid chromatography (UPLC) coupled with tandem mass spectrometry (LC-MS/MS) using nano-spray ionization. The nano-spray ionization experiments were performed using a TripleT of 5600 hybrid mass spectrometer (ABSCIEX) interfaced with nano-scale reversed-phase UPLC (Waters corporation nano ACQUITY) using a 20 cm-75 micron ID glass capillary packed with 2.5- μ m C18 (130) CSHTM beads (Waters corporation). Peptides were eluted from the C18 column into the mass spectrometer using a linear gradient (5–80%) of ACN (acetonitrile) at a flow rate of 250 μ l min⁻¹ for 1 h. The buffers used to create the ACN gradient were: buffer A (98% H₂O, 2% ACN, 0.1% formic acid, and 0.005% trifluoroacetic acid (TFA)) and buffer B (100% ACN, 0.1% formic acid, and 0.005% TFA). MS/MS data were acquired in a data-dependent manner in which the MS1 data was acquired for 250 ms at *m/z* of 400 to 1,250 Da and the MS/MS data was acquired from *m/z* of 50 to 2,000 Da. The independent data acquisition (IDA) parameters were as follows; MS1-TOF acquisition time of 250 ms, followed by 50 MS2 events of 48 ms acquisition time for each event. The threshold to trigger MS2 event was set to 150 counts when the ion had the charge state +2, +3 and +4. The ion exclusion time was set to 4 s. Finally, the collected data were analysed using Protein Pilot 4.5 (ABSCIEX) for peptide identifications.

Three-dimensional-immuno-FISH. Cells were grown on poly-lysine coated cover slips. For immuno-FISH cells were fixed with 4% formaldehyde and permeabilized with 0.5% Triton X-100 followed by glycerol wash and freezing in liquid nitrogen. Cells were then treated with 0.1 M HCl and with pepsin. Cells were then washed, blocked in 4% BSA and incubated with the primary anti-matrin-3 antibody (ab70336). The cover slips were incubated with a biotinylated secondary antibody (Vector Labs BA-1000) followed by fluorescent streptavidin and post-fixed cells at 4% paraformaldehyde. The cover slips were then immersed in 50%formamide/2XSSC. The DNA on the slides was denatured together with fluorescent probes for 10 min at 76 °C. The probes were prepared from BACS (purchased from Children's Hospital Oakland Research Institute) using a nick translation kit according to the manufacturer's instructions (Roche). Hybridization took place at 37 °C for at least 2 days in a humid chamber. After hybridization cover slips were washed in 50% formamide/2 \times SSC and in 0.5 \times SSC. Afterwards the cover slips were incubated with the appropriate fluorescent secondary antibodies, washed and mounted on slides with ProLong gold containing DAPI. The images were taken using an Olympus FV1000 confocal microscope.

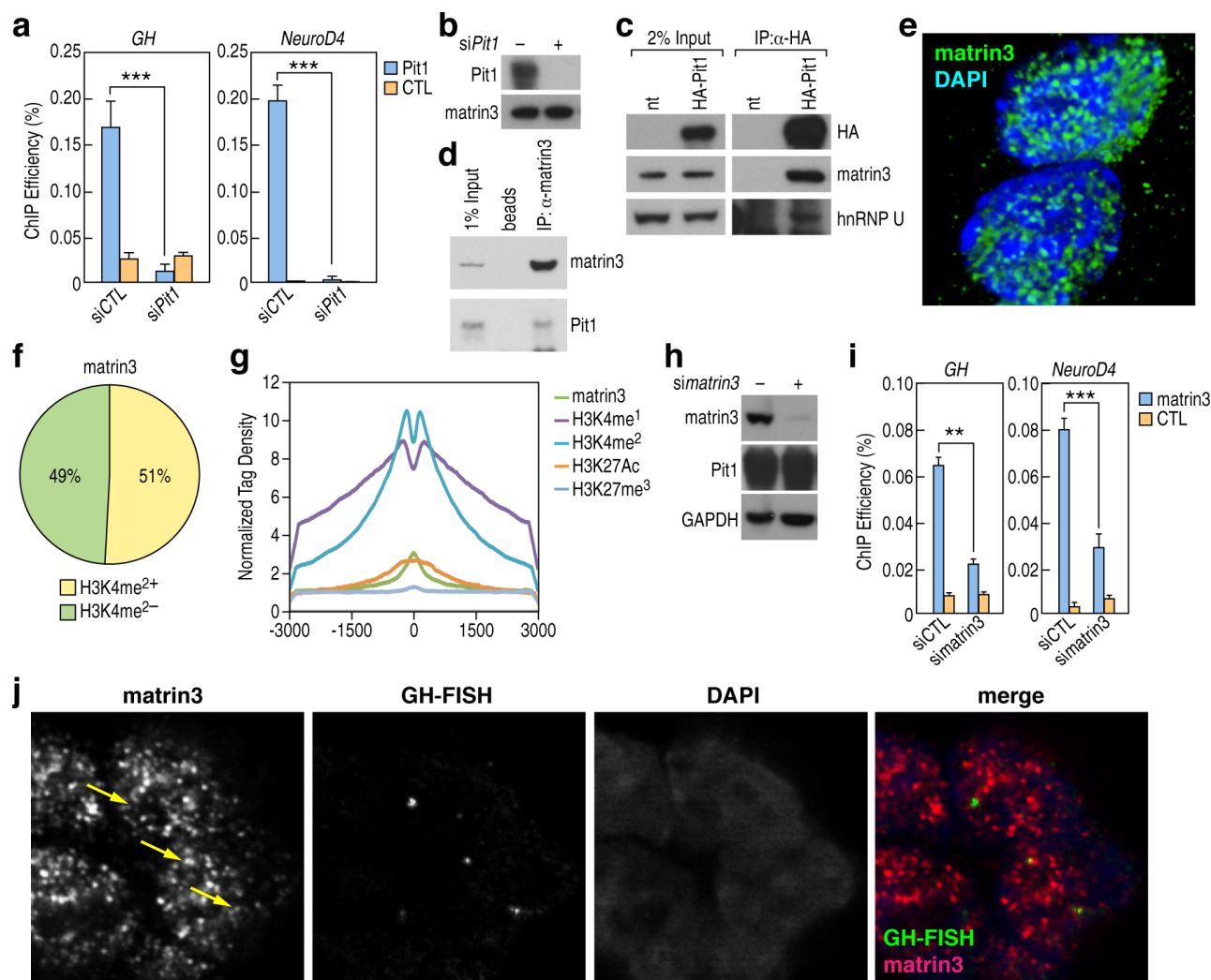
Three-dimensional-immuno-FISH data analysis. Quantification of co-localization was determined using the Volocity Software (Perkin Elmer) as a double blind experiment. The FISH signals and Matrin-3 spots were identified automatically based upon intensity and object size. The number of overlapping signals was automatically determined. The data are based on at least 200 FISH signal per condition from two independent experiments. Up to 10 pictures per conditions were analysed, *P* value was calculated using Student's two tailed *t*-test.

LIS extraction. Biochemical extraction of LIS resistant fraction was performed according to the published detailed protocol³³ with the following modifications: LIS extraction was optimized to be the most successful at 7.5 mM, DNase I digestion (5 U per 50 µl of nuclei) was performed on ice 15 min.

Mice. *Prop1*-Cre transgenic mice were generated by using the 2.2 kb promoter and enhancer region of the *Prop1* gene to drive the Cre recombinase expression in mixed

background (X. Zhu, M.G.R. *et al.*, submitted). *Satb1*^{fllox/fllox} mice, which have LoxP recombination sites flanking *Satb1* exons 3 and 5 has been generated in the T. Kohwi-Shigematsu laboratory (M.A. Balamotis, T.K.-S., K.Y. *et al.*, submitted). Mice were collected at embryonic day 16.5, WT = 4, cKO = 12 embryos.

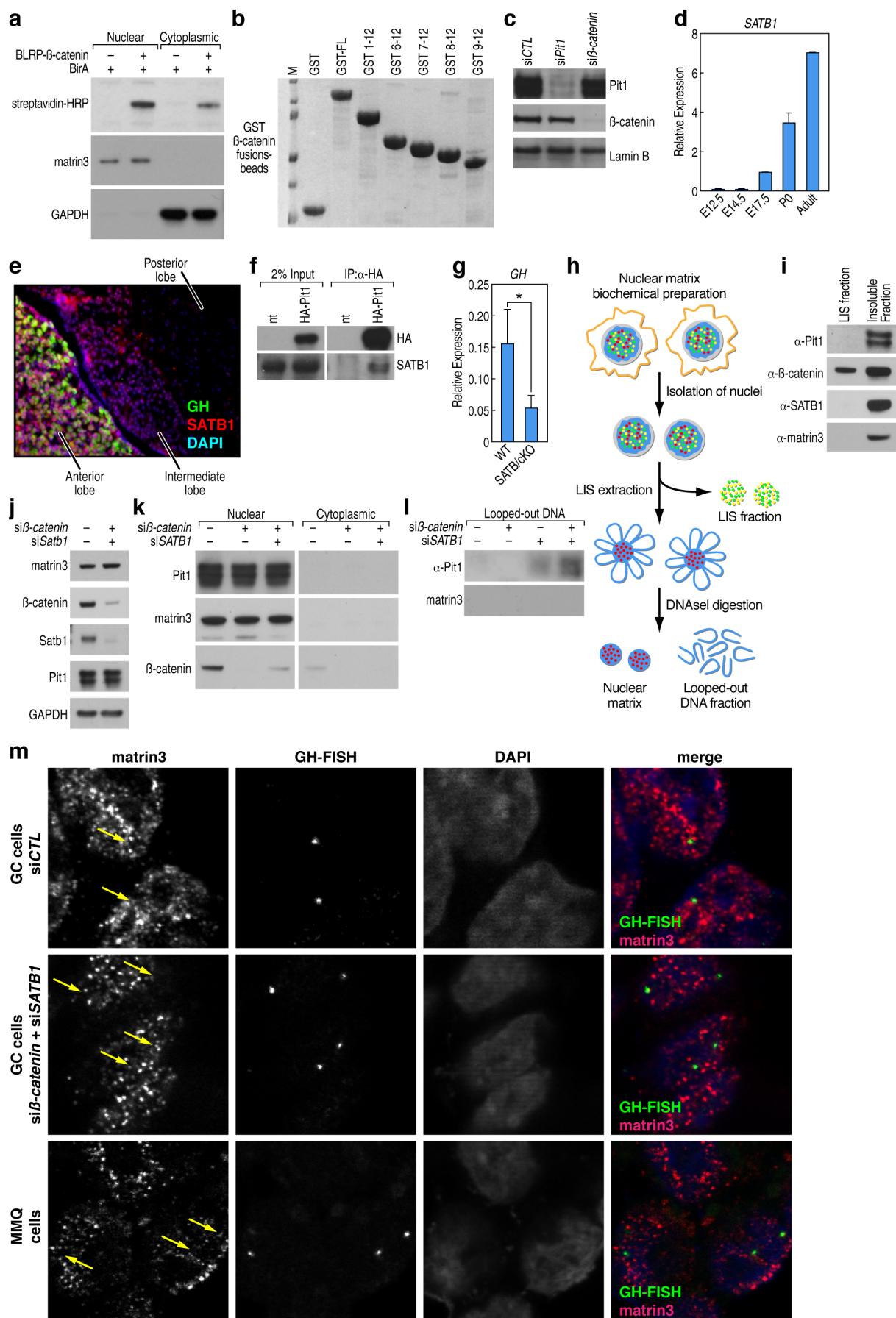
31. Wang, D. *et al.* Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* **474**, 390–394 (2011).
32. Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845–1848 (2008).
33. Gasser, S. M. & Vassetzky, Y. S. in *Chromatin: A Practical Approach* (ed. Gould, H.) pp. 111–124 (Oxford Univ. Press, 1998).



Extended Data Figure 1 | Matrin-3 co-localizes with enhancers in GC cells.

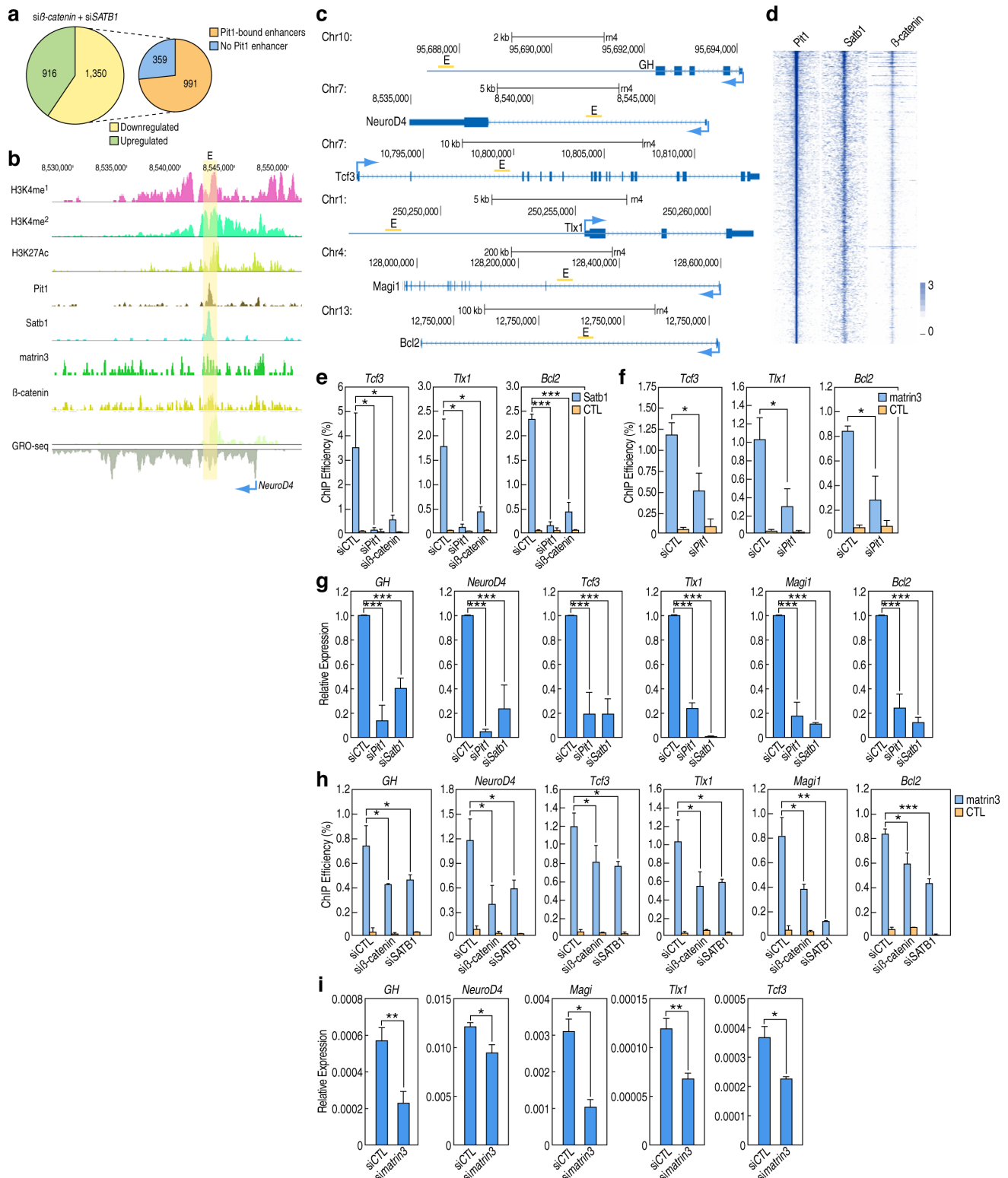
a, Validation of Pit1 antibody specificity by qPCR analysis of ChIP signals on two known targets; the GH and NeuroD4 enhancers shows lack of Pit1 signal after Pit1 knockdown. Experiment repeated 2 times, P values were calculated using Student's two tailed t -test (\pm s.d.; *** $P < 0.001$). **b**, Western blot confirming Pit1 knockdown efficiency in samples used to assess Pit1 antibody specificity. **c**, Co-immunoprecipitation of HA-tagged, overexpressed Pit1 protein from 293T cells confirms interaction of Pit1 protein with endogenous matrin-3 and hnRNP-U proteins (nt, non-transfected). **d**, Reciprocal co-immunoprecipitation of matrin-3 followed by western blot analysis confirmed interactions of matrin-3 with Pit1. **e**, Super-resolution image ($\times 100$, OMX DeltaVision) of immunostaining with anti-matrin-3 antibody in GC cells reveals "structure-like" matrin-3 network. **f**, ChIP-seq analysis reveals that $\sim 50\%$ of matrin-3 DNA binding sites co-localize with the

enhancer mark, H3K4me2, in GC cells. **g**, Meta-analysis plot of ChIP-seq data shows that Pit1 bound enhancers co-localize with matrin-3, H3K4me1, H3K4me2 and H3K27Ac, but not with a mark of silent chromatin, H3K27me3. x -axis is the distance from the Pit1 peak in base pairs. **h**, Western blot confirming matrin-3 knockdown efficiency in GC cells. **i**, Validation of matrin-3 antibody specificity by qPCR analysis of ChIP signal on known Pit1 targets shows lack of matrin-3 signal after matrin-3 knockdown. Experiment repeated 2 times, P values were calculated using Student's two tailed t -test (\pm s.d.; ** $P < 0.01$, *** $P < 0.001$). **j**, Single optical sections of immuno-FISH experiment in GC cells. From left: matrin-3, GH-FISH, DAPI, merge; each staining shown separately in black and white, and merged as matrin-3 antibody staining in red and the GH locus labelled with a DNA probe in green.



Extended Data Figure 2 | Co-localization and interaction of Pit-1 with β -catenin and Satb1. **a**, Western blot showing that biotinylated BLRP- β -catenin is present in the nuclear and cytoplasmic fractions of GC cells. Biotinylated BLRP- β -catenin is visualized with streptavidin-horse radish peroxidase. **b**, Bacterially expressed GST- β -catenin fusion proteins used to map interaction with Pit-1 in GC cell extracts. **c**, Western blot validation of siRNA knockdown of β -catenin and Pit-1 proteins in GC cell samples. **d**, RT-qPCR analysis of *Satb1* mRNA expression in embryonic and adult mouse pituitary glands showing increased expression that parallels differentiation of Pit1-dependent cell types, $n = 3$, \pm s.d. **e**, Immunohistochemical staining of SATB1 and GH in the anterior lobe of an adult mouse pituitary showing co-localization of signals at the cellular level. **f**, Western blot showing co-immunoprecipitation of SATB1 with HA-tagged Pit1 in 293T cells. **g**, RT-qPCR analysis of *GH* nascent transcript levels in E16.5 mouse pituitaries

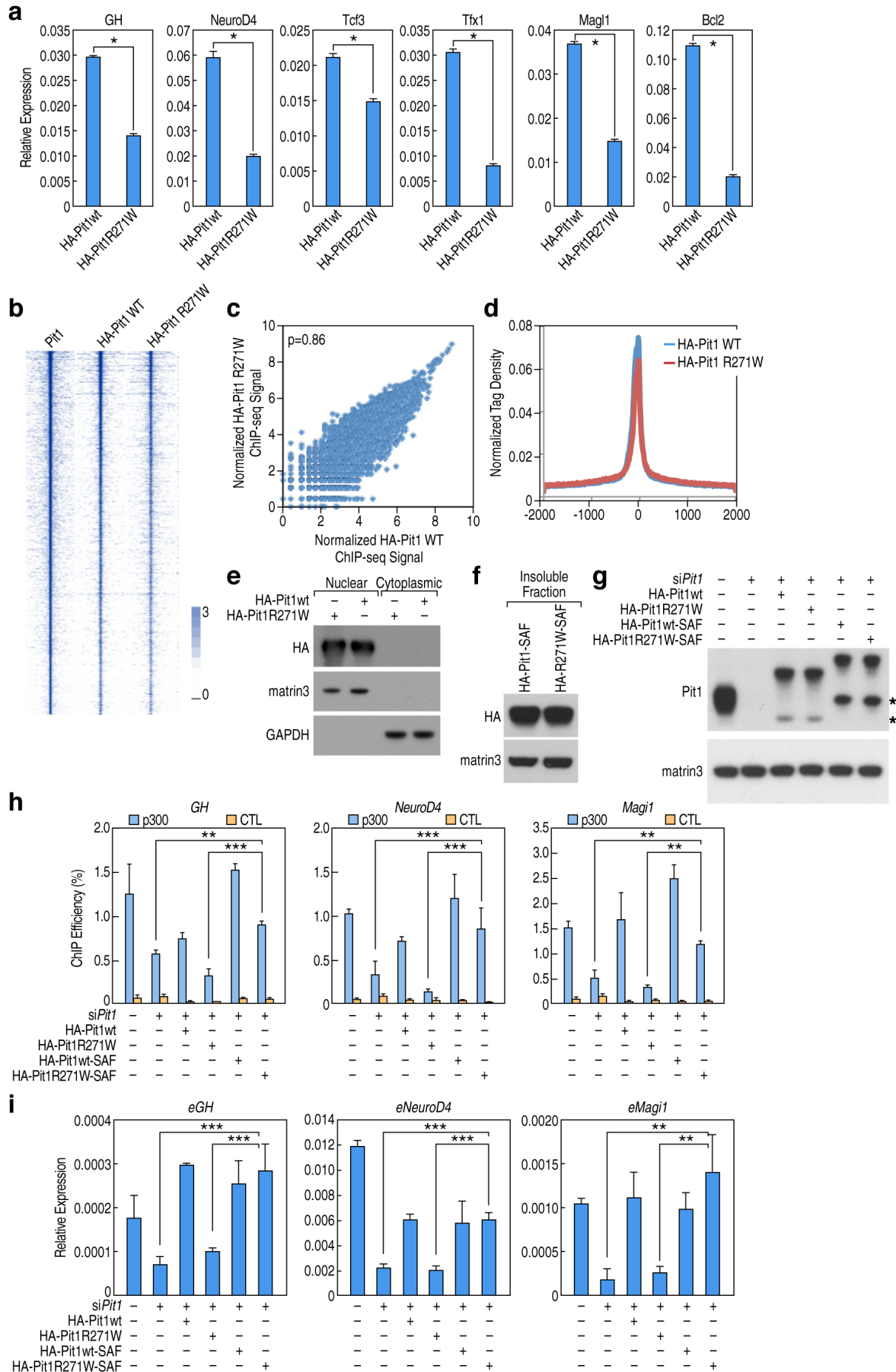
harbouring a Prop1-CRE conditional deletion of *Satb1*, \pm s.d., $*P < 0.05$, WT = 4, cKO = 12. **h**, Schematic representation of the LIS-extraction procedure used to isolate matrin-3-rich network. **i**, Pit-1 co-localizes with *Satb1* and matrin-3 in the LIS-resistant insoluble fraction whereas β -catenin is present in both LIS-extracted and LIS-resistant insoluble fractions in GC cells. **j**, Western blot confirming β -catenin and *Satb1* knockdown efficiency in GC cells. **k**, siRNA knockdown of β -catenin and *Satb1* in GC cells does not affect either the subcellular distribution or the level of Pit1 protein. **l**, After siRNA knockdown of β -catenin and *Satb1* followed by LIS extraction, Pit-1 is detectable in the “looped-out DNA” fraction. **m**, Immuno-FISH in GC cells before and after siRNA knockdown of β -catenin and *Satb1* and in MMQ cells. Single optical section of each experiment is presented. In merge: Matrin3 antibody immunostaining in red and GH loci labelled with a fluorescent green DNA probe.



Extended Data Figure 3 | β -catenin and SATB1 have transcriptional effects on Pit1 targets.

a, Most β -catenin/Satb1-activated target genes are associated with Pit-1-bound enhancers in GC cells. **b**, Example picture of ChIP-seq and GRO-seq analysis of a Pit-1/Satb1 target gene, the *NeuroD4* locus. **c**, Location of Pit1-dependent enhancers (in yellow) relative to the transcription start sites of selected target genes. **d**, Heat-map plot of ChIP-seq tag distribution on Pit1 enhancers showing enrichment of Satb1 and β -catenin signal in centre of enhancer. **e**, ChIP-qPCR analysis of Satb1 association with *Pit1* enhancers upon either Pit1 or β -catenin knockdown. Experiments repeated 3 times, *P* values calculated using Student's two tailed *t*-test (\pm s.d.; **P* < 0.05, ***P* < 0.01, ****P* < 0.001). **f**, ChIP-qPCR analysis of matrin-3 association with

Pit1 enhancers upon Pit1 knockdown. Experiments repeated 2-4 times, *P* values calculated using Student's two tailed *t*-test (\pm s.d.; **P* < 0.05, ***P* < 0.01, ****P* < 0.001). **g**, qPCR analysis showing significant change in expression of selected Pit1 target genes upon either Pit1 or Satb1 knockdown (\pm s.d.; ****P* < 0.001). **h**, ChIP-qPCR analysis of matrin-3 association with Pit1-dependent enhancers following siRNA knockdown of β -catenin or Satb1. Experiment repeated 2-3 times, *P* values were calculated using Student's two tailed *t*-test (\pm s.d.; **P* < 0.05, ***P* < 0.01, ****P* < 0.001). **i**, RT-qPCR analysis of relative expression of target gene mRNA following siRNA knockdown of matrin-3.



Extended Data Figure 4 | Pit-1 R271W mutant protein binds cognate DNA sites but does not associate with the nuclear LIS-resistant insoluble fraction.

a, RT-qPCR analysis of mRNA expression of Pit1-dependent target genes in GC cells. Both overexpressed, wild-type and R271W, Pit1 proteins compete for binding to recognition sites with endogenous Pit1. Experiment repeated 2 times, *P* values were calculated using Student's two tailed *t*-test (\pm s.d.; $*P < 0.05$). **b**, Heat map of ChIP-seq data showing that overexpressed HA-Pit1 WT and HA-Pit1 R271W bind to the same enhancer DNA recognition sites as endogenous Pit1 protein. **c**, Scattered dot plot of genome-wide peaks consistency analysis between the wild-type and R271W Pit1 to Pit1 enhancers with Pearson correlation coefficient shows a strong resemblance of the data. **d**, ChIP-seq analysis comparing normalized tag density on GC samples

containing overexpressed wild-type and R271W Pit1 protein. *x*-axis represents the number of base pairs from the centre of the Pit1 peak. **e**, Both wild-type and R271W Pit1 protein express at similar levels and partition in the nuclear fraction. **f**, Fusion of the SAF matrix-association domain results in fractionation of R271W-SAF Pit1 protein in LIS-resistant insoluble fraction of GC cells. **g**, Western blot showing equivalent expression of all HA-tagged Pit1 constructs in GC cells. Asterisk, product of degradation of overexpressed Pit1 protein. **h**, **i**, ChIP-qPCR (**h**) and RT-qPCR (**i**) analysis of effect of overexpression of different forms of Pit1 protein on p300 association and enhancer RNA (eRNA) expression, respectively, in the absence of endogenous Pit1. Experiments have been repeated 2–4 times, *P* values were calculated using Student's two tailed *t*-test (\pm s.d.; $**P < 0.01$, $***P < 0.001$).

CORRIGENDUM

doi:10.1038/nature13876

Corrigendum: A global strategy for road building

William F. Laurance, Gopalasamy Reuben Clements, Sean Sloan, Christine S. O'Connell, Nathan D. Mueller, Miriam Goosem, Oscar Venter, David P. Edwards, Ben Phalan, Andrew Balmford, Rodney Van Der Ree & Irene Burgues Arrea

Nature **513**, 229–232 (2014); doi:10.1038/nature13717

In this Letter, as a result of an inadvertent spreadsheet error, four values presented in Table 1 were slightly inflated. These relate to the proportions of Earth's total land surface located within the 'conserve', 'agriculture', 'conflict' and 'low-tension' zones. The correct percentage values for these four zones under the 'global' heading are 42.96, 11.40, 15.34 and 30.30, respectively. Two of these values (the global percentages for conflict and agriculture zones) were also mentioned in the main text. We apologise for these errors, which have now been corrected in the online versions of the Letter, and do not affect the interpretation of our analyses.

CAREERS

@NATUREJOBS Follow us on Twitter for the latest on jobs and careers go.nature.com/e492gf

NATUREJOBS BLOG The latest on careers news and tips <http://blogs.nature.com/naturejobs>

NATUREJOBS For the latest career listings and advice www.naturejobs.com



FAMILY

Community care

Help is available for researchers who want to provide care for elderly relatives.

BY HELEN SHEN

Every second year for almost ten years, cell biologist Susan Michaelis has looked forward to a trip to Innsbruck, Austria, for a conference that she helps to organize.

But in 2010, after months of planning and preparation, she had to cancel her travel plans to make an emergency trip to Florida. Her mother, who lived there, was suffering from excruciating joint pain due to osteoarthritis, and needed help. The crisis sparked her mother's ultimate decline

into immobility, and marked a turning point for Michaelis — the start of a journey into long-term caregiving. It has been tough. "I'm glad I didn't know what was coming and how much of me it was going to consume," she says. "It just felt great to be able to be there to help her."

Today, Michaelis, a professor at Johns Hopkins University in Baltimore, Maryland, is among the estimated 17% of full-time workers in the United States who juggle their careers with medical, financial or legal responsibilities for ageing parents. Globally, the problem

is expected to worsen: the World Health Organization estimates that nearly one in four people will be older than 60 by 2050.

Academics tend to have relatively flexible schedules, but basic research often leaves little time free for giving hours of care to an ageing or ill parent. "If this had happened to me at a point in my career ten years earlier, I couldn't have done what I did — because I would be in there bucking for tenure," says Michaelis.

Combining a demanding career with caregiving creates what seems like enough work for a small army. Accepting help is an important step towards balancing these demands. Experienced carers say that reaching out to family members, friends, colleagues and institutional assistance programmes has been invaluable.

LONG-DISTANCE RELATIONSHIP

Distance has emerged as a significant problem in caregiving, says Meg Stoltzfus of the Johns Hopkins Office of Work, Life and Engagement, where she coordinates support for faculty members and staff who care for elderly relatives. "By the time a faculty member has finished graduate school and found a job, they're unlikely to be in the same region as their parents," she says.

Johns Hopkins is one of several US institutions that employ work-life staff who specialize in care of elderly relatives — others include the University of California, Berkeley, and the University of Arizona in Tucson. These staffers can refer faculty members to care providers who can take a relative to medical appointments, help to prepare meals or provide full-time medical care.

Such care is not cheap. Costs vary widely by region and by level of care: in the United States, for example, an in-home health aide can cost around US\$30,000 a year; a room in a private nursing facility might be more than \$94,000 a year. In countries including the United States, France and Belgium, government social services provide subsidies for the care of low-income senior citizens. But for those of middle income, long-term care is a significant financial burden.

For years, Michaelis had flown to Florida every 4–6 weeks to help her mother for a few days at a time. Her father, who had handled the couple's finances, errands and household repairs, died in 2007 at the age of 97, leaving Michaelis's mother — who had stopped driving — ill-equipped to take care of herself. By 2010, she was increasingly unable to walk, and Michaelis needed outside help.

Michaelis's mother wanted to stay in her own home, and Michaelis decided to help to keep her there for as long as possible. Through ►

► a friend's referral, she hired daytime and overnight assistants to help her mother with meal preparation, trips to the bathroom and some shopping and housekeeping tasks.

Michaelis and her brother called their mother daily, and Michaelis continued to visit regularly. Much of her work, including grant applications and reviewing papers, could be transported easily. But over the years, Michaelis says, the time and energy spent arranging her mother's medical appointments, managing health-insurance paperwork and dealing with unexpected household and medical emergencies — often from afar — took attention away from her research programme.

"We all already have a lot of non-scientific work that we have to do. Even a small diminishment in time has an impact," says Michaelis.

This year, Michaelis noticed a decline in her mother's mental acuity, and in April, she moved her to an assisted-care facility in Baltimore. She had to delay renewing one of her grants so that she could help her mother to move and prepare her parents' house for sale.

"It has had an impact on the progress of our work," she says. But, she adds, "I would do it all over again".

RETURNING HOME

Faced with the worry and expense of providing care and oversight remotely, some researchers decide to relocate. That decision hinges on a combination of family and cultural expectations, as well as on personal goals: some researchers must decide whether to stay in an institution where their careers are established or move closer to their parents, where prospects might be uncertain.

Neuroscientist Fuqiang Xu grew up in a small farming town in Henan province, China. He moved to the United States in 1988, where he earned two doctorates, completed a postdoc and secured a research position at Yale

University's Magnetic Resonance Research Center in New Haven, Connecticut.

But in 2004, his mother died after a three-month battle with lung cancer, and he decided that he needed to move closer to his father, then in his eighties. "I should have returned earlier," says Xu. He accepted a research post in 2007 at the Wuhan Institute of Physics and Mathematics in China, about 400 kilometres from his hometown. He was able to visit several times before his father died in 2008.

Xu's colleagues often ask him if he regrets leaving the United States, but he says that he enjoys the lower competition for funding in China. And his prestigious US education and training, he says, have helped him to secure — and to support — his current position, where he heads a lab of about 40 postdocs, graduate students and other staff members.

COMPETING DEMANDS

Igor Stagljär, a molecular geneticist at the University of Toronto in Canada, is considering moving closer to his mother in Croatia, where she has been living alone since Stagljär's father died in 2007. Four years ago, Stagljär dropped a grant application and flew to Croatia for three weeks when he found that his mother had been diagnosed with renal cancer. "I was just mentally not ready to write that grant," he says.

Stagljär's mother made a full recovery. But the incident started him thinking about the future. "I'm the only child in the family," he says. "I would feel really bad if I couldn't help her."

He hopes to find a professorship in Europe once the younger of his two daughters, now 14, finishes secondary school in Canada. For now, he talks to his mother on Skype most days, and visits about three times a year, sometimes combining his trips with conference travel.

Caring for ailing parents while juggling busy work schedules can leave many researchers feeling overwhelmed. Adult children of parents

stricken with terminal illness or dementia can feel guilty, frustrated or grief-ridden.

In January 2012, as immunologist Laurie Glimcher started a job as dean of Weill Cornell Medical College in New York City, her father

"I'm the only child in the family. I would feel really bad if I couldn't help."

began showing signs of dementia. She moved him nearby, and tried to spend as much time as she could with him, even as she was learning

the ropes at her new position and managing a packed agenda of meetings and fund-raising events. But within a year, he lost interest in daily activities, including eating and drinking.

"It was a huge weight, watching him deteriorate," she says. Her father — Melvin Glimcher, former chair of orthopaedics at Massachusetts General Hospital in Boston — had published papers with her when he was well. "I wanted to remember him the way he was," she says.

In the months before her father's death in May, Glimcher found support in talking to colleagues who had lost parents to Alzheimer's disease. Some US universities offer counselling and guidance for carers, as well as free referrals for care providers, and free or subsidized back-up care for occasions when normal arrangements fall through (see 'Helping hands'). But many employees remain unaware of such services.

"I don't think people necessarily think, when they're having a personal crisis, to look to their employers first," says Ian Reynolds, president of the College and University Work-Life-Family Association, a nationwide organization of work-life professionals. "It's still a developing service and developing field," he says. But, he adds, "we're seeing a real need emerge".

In addition to institutional resources, many people rely on informal support. Michaelis chats to friends about her mother's care. Several have had their own care stories and tips, making more bearable what otherwise would have been a perplexing and lonely experience.

"There's kind of a trajectory you can expect for children, for their growth and development. Elder care is complicated. There are many different ways that people can fail, and it's totally unpredictable," says Michaelis. "It's really helpful to have a support group of friends to talk about all of this." ■

Helen Shen is a freelance writer living in Sunnyvale, California.

HELPING HANDS

Resources for caregivers

Caring for an ailing parent can be a daunting task. Here are some places to find help.

University work-life office Many institutions offer referrals for full- or part-time professional caregivers, who may have medical training. In case of brief, unexpected interruptions in care, some universities offer free or subsidized back-up care services.

Government agencies The US government funds a network of agencies that can help to find local programmes such as adult day-care and meal services. Families may be able to find help with health insurance and legal issues (see go.nature.com/nfhxpk). Countries

including Germany, Japan, the Netherlands and South Korea run national insurance systems that help to offset the costs of care.

Geriatric care-management services Care managers can help to develop and execute a care plan and perform in-home assessments to determine a care recipient's medical and social needs (www.caremanager.org).

Patient-advocacy groups Advocacy groups that focus on a particular disease, such as the Alzheimer's Association in Chicago, Illinois (www.alz.org), can connect patients and families to others who have gone through similar experiences. **H.S.**

CORRECTION

In the Careers Feature 'Kid-friendly digs' (*Nature* **513**, 575–577; 2014), the image including Suzanne Anderson had an incorrect caption. It should have read: 'Suzanne Anderson (far right) sometimes mixes family and research on trips to study glacier outbursts.'

USEFUL OBJECTS

It's time to play your part.

BY ERICA L. SATIFKA

After he passes the age of reason, my brother chooses to become a foundation. Specifically, the foundation of the new state capitol building in Austin, Texas.

"You've never been to Texas," I say.

"It was the best opening they had," he replies with a small, sad shrug. "And you get weekends off."

I'm still working the counter at Jiffy Mart, delaying the inevitable at a pointless task nobody asked me to perform. My friends have all gone off to be fire hydrants or ATMs or jackhammers or five-piece dinette sets. "Undifferentiated," the Makers call shirkers like me. I hear them whispering through the thought rays that emanate from their human-powered satellites *Choose. Decide. Be of use.*

And I reply: "Not yet."

The bell jingles and I look up to see a woman in mid-transformation barging into the store. Probably just took her injection after a weekend of rest. She's half-human, half-Vespa, and her chassis scrapes the paint off the door frame.

"A little help?"

I sigh and manoeuvre her through. I would have brought her purchases out to the parking lot if she'd asked me to. "What do you need?"

"Motor oil. Oreos."

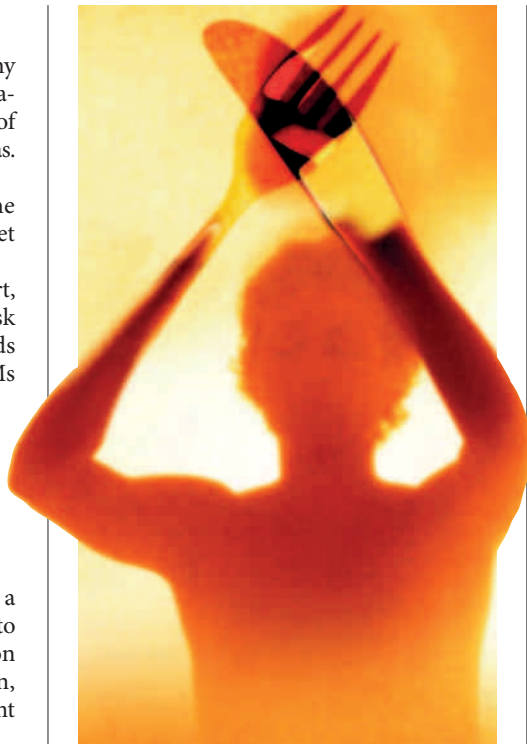
I tuck the items into her saddlebags. No charge, of course. As I close the door behind her, she belches a cloud of exhaust into my face. The transformation complete, she idles at the corner until a passing Maker hops aboard. It pops an Oreo into its mouth and speeds off, jagged teeth covered in chocolate bits.

The Makers are alive, but they're not organic. The division between 'living thing' and 'object' doesn't exist for them. And they have a hard time believing that we care about such a piddling thing as keeping our own bodies. To a Maker, a job's a job, and we all play our role. Except for us selfish undifferentiated types.

Choose, say the voices in my head. Decide. "Not yet," I say. "Piss off."

That keeps the voices down, for a little while anyway.

The Makers' home world is as artificial as they are, a spherical factory orbiting a distant blue sun. No nature, just industry. They arrived in the bodies of the last race they'd conquered, ships that died on contact with



our atmosphere. The ships died happy, the Makers told us, knowing they had been of use.

I'm not so sure about that.

It was a slow invasion masked as a self-improvement regimen. None of my friends really had a job. We were all living on plastic, taking useless classes at the community college to maintain our health insurance while we pretended our parents' basements were fabulous studio apartments. The queues for the Makers' employment centres stretched down the sidewalk like an ant trail.

Except, I kind of liked the art-history class I was enrolled in at the time. I didn't mind living at home. And anyway, injections hurt.

The Makers tend their human machines like careful gardeners. They shamle down the human-lined streets on their twisted, insectoid legs. And every day I feel their alien hate pulsing at me.

When they were full-time people, my parents used to telecommute, so it made sense for them to become a house. Which is great in one way, because I don't have to sleep

in a stranger's arm-pit. But it's also bad in another, because there isn't any privacy. Sex becomes

unthinkable in a house built from your parents' bones. I think that's why my brother's moving so far away.

"I'll miss you," I say as I watch my brother pack his bags.

He's already taken his starter injections, and his words come out thick and gravelly. Stone man. "You could come with me."

"I don't know anyone down there," I say. I don't add that I don't know anyone here anymore either. We don't wear nametags or anything, so you only know your friends when they transform into their part-time human forms. And useful objects don't want to hang around with undifferentiated slackers like me. "I'll write you. You'll still be able to read, won't you?"

"Of course I'll be able to read. I can do anything you can do. Except move."

"But you get weekends off."

"Weekends," he says, "and alternate Wednesdays."

I don't take public transportation anymore, and I don't dare climb in a taxi, not when I could be entering the cab of my hated fifth-grade teacher. Luckily, I only live ten miles from the ocean. I grab my trusty bike, which was never alive, and pedal down the road to the coast.

Choose, the voices say. Contribute. Be of use. I pedal faster.

Because it's November, the ocean is deserted. I take off my shoes, roll up my trouser legs, and wade into the brackish water.

I choose to be the air, I think. I choose to be the rain on my face and the rocks beneath my feet, the waves crashing over the rocks and the sun beating down on the waves. I wish to disappear into nature, into Earth itself. That's something the Makers can't give us, for these things have no function. They are not of use.

I stand in the ocean until the pounding rain becomes too much to bear. My teeth chatter. But I just can't bring myself to leave. The rain drowns out the voices, and the dark keeps me from seeing the boats in the distance and wondering who they are, if they're anyone at all.

Someday, I know I'll have to choose. I can't remain undifferentiated forever.

Not yet, though, not yet. I'm not nearly ready yet. ■

Erica L. Satifka's fiction has previously appeared in *Clarkesworld* and *Daily Science Fiction*, among others. She lives in Portland, Oregon. Visit her online at www.ericasatifka.com.

➔ NATURE.COM
Follow Futures:
@NatureFutures
go.nature.com/mtoodm